FEDERAL INSTITUTE OF EDUCATION, SCIENCE AND TECHNOLOGY OF SANTA CATARINA – FLORIANÓPOLIS CAMPUS
ACADEMIC DEPARTMENT OF METAL-MECHANICS
BACHELOR'S DEGREE PROGRAM IN MECHATRONICS ENGINEERING

FILIPE BRAGA PIRES

Convolutional Neural Networks for Mass Flow Estimation and Powder Feed Monitoring in the EHLA Process

AACHEN, 2025.

FEDERAL INSTITUTE OF EDUCATION, SCIENCE AND TECHNOLOGY OF SANTA CATARINA – FLORIANÓPOLIS CAMPUS
ACADEMIC DEPARTMENT OF METAL-MECHANICS
BACHELOR'S DEGREE PROGRAM IN MECHATRONICS ENGINEERING

FILIPE BRAGA PIRES

Convolutional Neural Networks for Mass Flow Estimation and Powder Feed Monitoring in the EHLA Process

Bachelor's Thesis submitted to the Federal Institute of Education, Science and Technology of Santa Catarina as part of the requirements for obtaining the degree of Mechatronics Engineer.

Advisor:
Prof. Cassiano Bonin, M. Eng.

Co-advisor:
Max Gero Zimmermann, M. Sc.

AACHEN, 2025.

Ficha de identificação da obra elaborada pelo autor.

Pires, Filipe
Convolutional Neural Networks for Mass Flow Estimation and Powder Feed Monitoring in the EHLA Process / Filipe Pires; orientação de Cassiano Bonin; coorientação de Max Gero Zimmermann. - Florianópolis, SC, 2025.
87 p.

Trabalho de Conclusão de Curso (TCC) - Instituto Federal de Santa Catarina, Câmpus Florianópolis. Bacharelado em Engenharia Mecatrônica. Departamento Acadêmico de Metal Mecânica.
Inclui Referências.

1. Manufatura aditiva. 2. EHLA. 3. Aprendizado de máquina. 4. Monitoramento em tempo real. 5. Redes neurais convolucionais. I. Bonin, Cassiano. II. Gero Zimmermann, Max. III. Instituto Federal de Santa Catarina.
IV. Convolutional Neural Networks for Mass Flow Estimation and Powder Feed Monitoring in the EHLA Process.

Convolutional Neural Networks for Mass Flow Estimation and Powder Feed Monitoring in the EHLA Process

FILIPE BRAGA PIRES

This thesis has been deemed suitable for the fulfilment of the requirements for the bachelor's degree in Mechatronics Engineering and was approved in its final form by the evaluation committee of the Mechatronics Engineering Program at the Federal Institute of Education, Science and Technology of Santa Catarina.

Aachen, July 25th, 2025

Evaluation Committee Members:

gov.br Documento assinado digitalmente
CASSIANO BONIN
Data: 20/08/2025 17:20:57-0300
Verifique em https://validar.iti.gov.br

Prof. Cassiano Bonin, M. Eng.

IFSC Florianópolis

gov.br Documento assinado digitalmente
MAURICIO EDGAR STIVANELLO
Data: 20/08/2025 17:35:48-0300
Verifique em https://validar.iti.gov.br

Prof. Maurício Edgar Stivanello, Dr. Eng.

IFSC Florianópolis

Andres Gasser
Digital unterschrieben von Andres Gasser
Datum: 2025.08.21 08:52:44 +02'00'

Andres Gasser, Dr. Ing.

Fraunhofer ILT

To my parents,
whose strength, encouragement,
and unwavering belief in me have
guided me every step of the way.

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to my parents for their constant support and encouragement throughout my academic journey. I am also sincerely thankful to my advisor, Prof. Cassiano Bonin, M. Eng., for his guidance and the valuable insights provided during the development of this work. My appreciation goes as well to my supervisor here in Germany, Max Gero Zimmermann, M. Sc., for closely accompanying the project, offering practical support, and contributing significantly to the progress of this thesis. I would like to thank the Fraunhofer Institute for Laser Technology (ILT) for granting me the opportunity to conduct this research with access to excellent infrastructure and resources. Finally, my special thanks to Prof. Erwin Werner Teichmann, Dr. Eng., who made this international exchange possible and helped establish the collaboration between institutions.

"If I have seen further, it is by standing on the shoulders of giants."
— Isaac Newton

ABSTRACT

Laser Material Deposition (LMD) and its high-speed variant, EHLA, are modern techniques used to enhance surface coating performance and repair metallic components. EHLA stands out for its ability to deposit thin, metallurgically bonded layers with minimal thermal impact, enabling rapid and precise coating of large areas. However, inconsistencies in powder feeding, caused by factors such as nozzle clogging, powder agglomeration, or environmental variations, can lead to defects, material waste, and system downtime. Conventional monitoring systems often lack the responsiveness and precision required for real-time anomaly detection. This thesis proposes a machine learning-based monitoring system designed to estimate powder mass flow during the EHLA process. The system leverages a convolutional neural network (CNN) based on EfficientNet-B1 encoders, trained to predict the disk speed of the powder feeder, which is experimentally correlated with the powder mass flow through weight measurements obtained using a precision balance. The model receives paired images from coaxial and side-view cameras, along with carrier gas flow rate data, as inputs. By learning the complex patterns associated with the visual and numerical data, the model can infer disk speed and thereby provide an indirect but effective estimate of powder flow behavior. Deviations between predicted and actual disk speed values are used to identify feeding anomalies in real time, allowing early intervention to maintain process stability. The best model was trained over the 4–16% disk speed range and achieved a Mean Absolute Error (MAE) of 0.257 in z-score units (approximately 0.12% disk speed error) and a coefficient of determination (R^2) of 0.9892 on the test set, while maintaining real-time prediction capability with an average inference time of 7 ms per sample. Comparative tests showed that side-view images alone yielded the best performance (MAE ~0.06%), and Grad-CAM visualizations confirmed the model's focus on physically meaningful image regions such as the powder stream and melt pool. This approach enables not only predictive monitoring but also provides real-time information to support operational decision-making. The proposed system offers a scalable and intelligent solution for enhancing process reliability in EHLA, aligning with the goals of Industry 4.0 by promoting automation, sustainability, and high-efficiency manufacturing.

Keywords: Additive Manufacturing. EHLA. Machine Learning. Real-Time Monitoring. Convolutional Neural Networks.

RESUMO

A Deposição a Laser de Material (LMD) e sua variante de alta velocidade, a Deposição a Laser de Material em Velocidade Extrema (EHLA), são processos avançados desenvolvidos para melhorar a eficiência e a qualidade de revestimentos superficiais e reparos de componentes. A EHLA, em particular, permite a aplicação de camadas metalurgicamente unidas com alta precisão e baixo aporte térmico, sendo adequada para revestir grandes superfícies de forma rápida e eficiente. No entanto, as inconsistências na alimentação de pó, causadas por fatores como entupimento do bico, aglomeração de pó ou variações ambientais, podem levar a defeitos, desperdício de material e tempo de inatividade do sistema. Os sistemas de monitoramento convencionais geralmente não têm a capacidade de resposta e a precisão necessárias para a detecção de anomalias em tempo real. Esta tese propõe um sistema de monitoramento baseado em aprendizado de máquina projetado para estimar o fluxo de massa de pó durante o processo de EHLA. O sistema aproveita uma rede neural convolucional (CNN) baseada em *encoders* EfficientNet-B1, treinada para prever a velocidade do disco do alimentador de pó, que é experimentalmente correlacionada com o fluxo de massa de pó por meio de medições de peso obtidas usando uma balança de precisão. O modelo recebe imagens sincronizadas de câmeras coaxiais e de visão lateral, juntamente com dados de taxa de fluxo de gás de transporte, como entradas. Ao aprender os padrões complexos associados aos dados visuais e numéricos, o modelo pode inferir a velocidade do disco e, assim, fornecer uma estimativa indireta, mas eficaz, do comportamento do fluxo de pó. Os desvios entre os valores previstos e reais da velocidade do disco são usados para identificar anomalias de alimentação em tempo real, permitindo a intervenção precoce para manter a estabilidade do processo. O melhor modelo foi treinado na faixa de velocidade do disco de 4% a 16% e obteve um erro médio absoluto (MAE) de 0,257 em *z-score* (aproximadamente 0,12% de erro na velocidade do disco) e um coeficiente de determinação (R^2) de 0,9892 no conjunto de teste, mantendo a capacidade de previsão em tempo real com um tempo médio de inferência de 7 ms por amostra. Testes comparativos mostraram que as imagens de visão lateral sozinhas produziram o melhor desempenho (MAE ~0,06%), e as visualizações do Grad-CAM confirmaram o foco do modelo em regiões de imagem fisicamente significativas, como o fluxo de pó e a poça de fusão. Essa abordagem permite não apenas o monitoramento preditivo, mas também fornece informações em tempo real para a tomada de decisão operacional. O sistema proposto oferece uma solução inteligente e escalável para aumentar a confiabilidade do processo EHLA, alinhando-se às metas da Indústria 4.0 ao promover a automação, a sustentabilidade e a fabricação de alta eficiência.

Palavras-chave: Manufatura aditiva. EHLA. Aprendizado de máquina. Monitoramento em tempo real. Redes neurais convolucionais.

LIST OF FIGURES

Figure 1 – View of the nozzle showing signs of clogging at the exit. Such obstructions can compromise powder flow and deposition stability. 18
Figure 2 – Conventional LMD 21
Figure 3 – Extreme High-Speed LMD 22
Figure 4 – Basic Artificial Neural Network (ANN) architecture with three distinct layers 26
Figure 5 – Gradient descent descending a loss landscape toward a local minimum 28
Figure 6 – A simple CNN architecture, comprised of just five layers 30
Figure 7 – A visual representation of a convolutional layer 30
Figure 8 – Grad-CAM class-discriminative localization. (a) Original image. (b) Grad-CAM heatmap for the class "dog". (c) Grad-CAM heatmap for the class "cat". 32
Figure 9 – EHLA Setup 34
Figure 10 – Machine setup used for data acquisition 35
Figure 11 – Reference curve: disk speed vs. measured powder mass flow 37
Figure 12 – Week 1 - with illumination and filters: (a) Side View and (b) Coaxial View 38
Figure 13 – Week 1 - without illumination and filters: (a) Side View and (b) Coaxial View 38
Figure 14 – Tube Week 1 - Lack of fusion 39
Figure 15 – Week 2: (a) Side View and (b) Coaxial View 40
Figure 16 – Particle Dispersion Increases with Carrier Gas Flow - Illustrative representation showing how particle dispersion varies with carrier gas flow. Distances d_1 and d_2 represent the spacing between particles under different flow conditions, with $d_2 > d_1$. 40
Figure 17 – Impact of Carrier Gas Flow on Powder Jet Behavior (10% Disk Speed) 41
Figure 18 – Tube produced during Week 2 using optimized parameters. Unlike earlier trials in Week 1, this deposition shows improved layer fusion and consistent material buildup. 41
Figure 19 – Week 2: (a) Side View ROI and (b) Coaxial View ROI 42
Figure 20 – Model architecture 43
Figure 21 – Training metrics over epochs for disk speed prediction: (a) MAE, (b) MSE, (c) RMSE, and (d) R^2. 52
Figure 22 – Validation metrics over training epochs for disk speed prediction: (a) MAE, (b) MSE, (c) RMSE, and (d) R^2. 53
Figure 23 – Training vs Validation Loss 54
Figure 24 – Evaluation metrics on the test set across training epochs: (a) MAE, (b) MSE, (c) RMSE, and (d) R^2. 55

Figure 25 – Predictions vs Actual 56
Figure 26 – Residuals Plot 57
Figure 27 – Residuals Histogram 58
Figure 28 – Quantile-Quantile plot of Residuals 59
Figure 29 – Predictions vs Actual for Step 10 (Interpolation Test) 60
Figure 30 – Residuals Plot for Interpolation Test (All Steps) 61
Figure 31 – Residuals Histogram for Step 10 Interpolation 62
Figure 32 – Quantile-Quantile plot of Residuals for Step 10 Interpolation 63
Figure 33 – Predictions vs Actual 64
Figure 34 – Residuals Plot 65
Figure 35 – Residuals Histogram 66
Figure 36 – Quantile-Quantile plot 67
Figure 37 – Predicted vs. Actual Plot 68
Figure 38 – Residuals Plot 69
Figure 39 – Residuals Histogram 70
Figure 40 – Q-Q Plot 71
Figure 41 – Grad-Cam Heatmap - Feature 8 73
Figure 42 – Grad-Cam Heatmap - Feature 7 73
Figure 43 – Grad-Cam Heatmap - Feature 6 74
Figure 44 – Spectral response curve of the Sony CMOS sensors used in both Basler camera models, highlighting peak sensitivity near 600 nm. . . 84
Figure 45 – Illustration of radiation laws 84
Figure 46 – Grad-Cam Heatmap - Feature 5 85
Figure 47 – Grad-Cam Heatmap - Feature 4 85
Figure 48 – Grad-Cam Heatmap - Feature 3 86
Figure 49 – Grad-Cam Heatmap - Feature 2 86
Figure 50 – Grad-Cam Heatmap - Feature 1 87
Figure 51 – Grad-Cam Heatmap - Feature 0 87

LIST OF TABLES

Table 1 – Powder Mass Flow Measurements 36
Table 2 – Model Evaluation Based on Input Combinations: MAE and Prediction Time Across Five Model Variants 72
Table 3 – Week 1 – Variation of Disk Speed and Carrier Gas 83
Table 4 – Week 1 – Fixed Parameters 83
Table 5 – Week 2 – Smaller Range (4 -16%) - Variation of Disk Speed and Carrier Gas 83
Table 6 – Week 2 – Fixed Parameters 83

LIST OF ABBREVIATIONS AND ACRONYMS

3D	Three-dimensional
AI	Artificial Intelligence
AM	Additive Manufacturing
ANNs	Artificial Neural Networks
ASTM	American Society for Testing and Materials
Adam	Adaptive Moment Estimation
AdamW	Adam with Weight Decay
CNC	Computer Numerical Control
CNNs	Convolutional Neural Networks
CWL	Central Wavelength
Concat	Concatenation
DED	Directed Energy Deposition
EHLA	Extreme High-Speed Laser Material Deposition
FWHM	Full Width at Half Maximum
GPU	Graphics Processing Unit
Grad-CAM	Gradient-weighted Class Activation Mapping
HAZ	Heat-Affected Zone
IFSC	Federal Institute of Santa Catarina
LED	Light Emitting Diode
LMD	Laser Material Deposition
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
NLPM	Normal Liters per Minute
R^2	Coefficient of Determination

RAM	Random Access Memory
RGB	Red, Green, Blue
RMSE	Root Mean Squared Error
RMSProp	Root Mean Square Propagation
ROI	Region of Interest
RWTH	Rhenish-Westphalian Technical University
ReLU	Rectified Linear Unit
SGD	Stochastic Gradient Descent
TabNet	Tabular Neural Network
UDP	User Datagram Protocol
VRAM	Video Random Access Memory
ViT	Vision Transformer
YAML	YAML Ain't Markup Language

CONTENTS

1 INTRODUCTION . . . 16
1.1 Justifications . . . 16
1.2 Problem Definition . . . 17
1.3 General Objective . . . 18
1.4 Specific Objectives . . . 18
2 STATE OF THE ART . . . 20
2.1 Process . . . 20
2.1.1 Directed Energy Deposition . . . 20
2.1.2 Laser Material Deposition . . . 20
2.1.3 Extreme High-Speed Laser Material Deposition . . . 21
2.2 In-situ Monitoring and Data-driven Modeling in Laser Deposition Processes . . . 22
2.3 Machine Learning and Neural Networks . . . 24
2.3.1 Overview of Machine Learning . . . 24
2.3.2 Supervised Learning . . . 25
2.3.3 Neural Networks . . . 25
2.3.4 Fundamentals of Learning in Neural Networks . . . 26
2.3.4.1 Gradient Descent and Backpropagation . . . 27
2.3.5 Convolutional Neural Networks (CNNs) . . . 29
2.4 Gradient-weighted Class Activation Mapping (Grad-CAM) . . . 31
2.5 Ludwig AI . . . 32
3 METHODOLOGY . . . 33
3.1 Methodology Overview . . . 33
3.2 Experimental Environment . . . 34
3.2.1 Sensor Positioning . . . 35
3.3 Data Acquisition and Preparation . . . 36
3.3.1 Experimental Setup and Materials . . . 36
3.3.2 Powder Mass Flow Measurement . . . 36
3.3.3 Experimental Protocol and Parameter Selection . . . 37
3.3.3.1 Week 1 – Exploratory Phase . . . 37
3.3.3.2 Week 2 – Optimized Setup (Final Dataset) . . . 39
3.3.4 Dataset Construction and Labeling . . . 41
3.4 Model Development and Integration . . . 42
3.4.1 Architecture Design . . . 43
3.4.2 Configuration File (YAML) . . . 45
3.4.3 Training Procedure . . . 48
3.4.4 Model Evaluation . . . 49
4 RESULTS . . . 50
4.1 Trained Models and Configuration . . . 50
4.2 Quantitative Evaluation . . . 51
4.2.1 Training and Validation Metrics . . . 51
4.2.2 Performance on the Test Set . . . 54
4.3 Visual Evaluation of Predictions . . . 56
4.3.1 Interpolation Evaluation: Step 10 Exclusion . . . 59
4.3.1.1 Predicted vs. Actual Plot . . . 60

4.3.1.2 *Residuals Plot* *61*
4.3.1.3 *Residuals Histogram* *61*
4.3.1.4 *Q-Q Plot of Residuals* *62*
4.3.2 Extra Intermediate points (Week 2 + Week 3) 63
4.3.3 Interpolation Behavior – Excluding Points 6, 10, and 14 67
4.4 Model Comparison **71**
4.5 Interpretability Analysis with Grad-CAM **72**
4.6 Final Considerations **75**
5 CONCLUSION AND FUTURE WORK **76**
5.1 Conclusion **76**
5.2 Future Work **77**
BIBLIOGRAPHY **79**
ANNEX **82**

1 INTRODUCTION

Laser Material Deposition (LMD) and Extreme High-Speed Laser Material Deposition (EHLA) are advanced manufacturing processes developed to increase the efficiency and quality of surface coatings and component repairs. EHLA, in particular, enables the application of metallurgically bonded layers with high precision and minimal heat input, making it suitable for coating large surfaces quickly and efficiently. These processes have shown significant industrial relevance, including applications in corrosion and wear protection, and the repair of heat-sensitive components (SCHOPPHOVEN; GASSER; BACKES, 2017). Despite their benefits, these processes are highly sensitive to variations in powder flow consistency, which can result in defects, increased material waste, and costly equipment downtime (BONIN *et al.*, 2023). These challenges are often intensified by process instabilities and inconsistencies in powder delivery, which remain difficult to monitor using conventional systems. (BONIN *et al.*, 2023).

The integration of machine learning (ML) into powder-based manufacturing processes offers transformative potential to address these issues. By leveraging real-time data from visual sensors and numeric process parameters, ML-based systems enable predictive maintenance and anomaly detection. Bonin *et al.* (2023) demonstrated that such systems could accurately predict clogging events and powder flow variations, reducing downtime and enhancing process reliability.

Aligning with the principles of Industry 4.0, which emphasize automation and data-driven decision-making, these advancements represent a significant step forward in manufacturing sustainability and efficiency. Integrating ML-driven sensors with traditional powder delivery systems not only improves quality and reduces waste but also ensures adaptability to dynamic manufacturing conditions. This research builds on these innovations to design an intelligent sensor system that integrates multi-modal data for real-time anomaly detection, laying the groundwork for smarter and more sustainable industrial applications.

1.1 Justifications

In directed energy deposition processes, such as LMD and EHLA, ensuring a consistent mass flow is crucial to achieving high-quality results. Variations in powder behavior caused by clogging, leakage, or uneven distribution can lead to defects, material waste, increased production costs, and equipment downtime. Current monitoring systems lack the precision and real-time capabilities needed to detect these issues before they impact the process. (BONIN *et al.*, 2023)

The aim of this work is to develop an intelligent sensor system that integrates cameras, numerical features, and machine learning to predict variations in mass flow in

real time. The synchronization of numerical data, such as flow rates, with high-speed camera images enables advanced analyses. Preprocessing techniques, including data normalization and visual feature extraction (such as edge detection and texture analysis), ensure compatibility between modalities.

Machine learning models, including convolutional neural networks (CNNs) for image-based inputs and fully connected layers for numerical features, are employed to capture patterns indicative of mass flow anomalies. The models are evaluated using standard regression metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2), ensuring robust assessment of predictive performance.

This approach allows for early detection of issues such as clogging or uneven distribution by combining visual information with numerical trends to recognize subtle patterns that would otherwise be overlooked. The robust integration of multimodal data, together with advanced analytical techniques, provides an effective solution for real-time monitoring and optimization of the process. Additionally, the system contributes to sustainability by reducing waste and establishes a new paradigm for applying machine learning in predictive manufacturing.

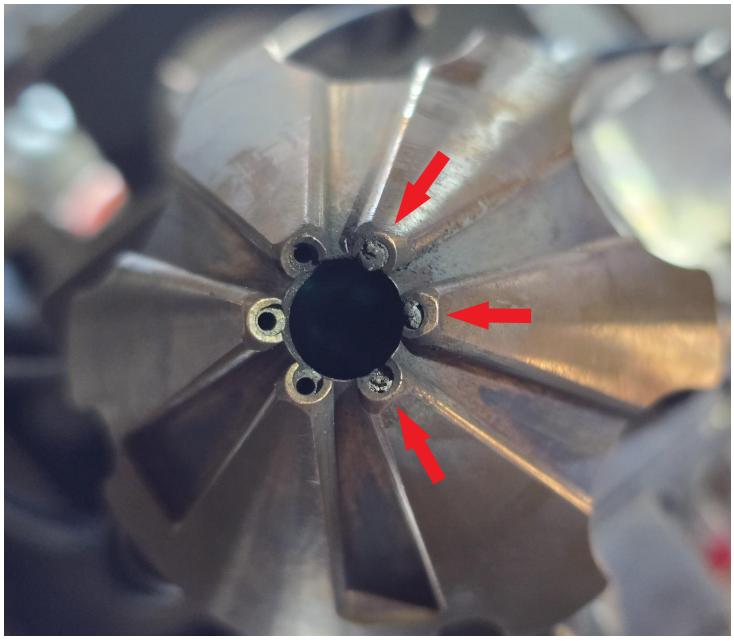
1.2 Problem Definition

Despite the critical role of powder mass flow in processes like EHLA and LMD, current monitoring methods remain inadequate for detecting early signs of disruption during operation. Many of these systems lack the sensitivity, temporal resolution, or integration with other process variables required for effective real-time diagnostics. As a result, subtle variations in powder delivery, caused by factors like clogging, environmental instability, or non-uniform distribution, often go unnoticed until visible defects or system failures occur.

Traditional monitoring approaches tend to rely on binary fault detection or isolated sensor readings, which limits their ability to anticipate process deviations. As highlighted by Bonin *et al.* (2023), these systems are not capable of predictive monitoring and thus fail to proactively prevent quality losses. There is a clear need for intelligent solutions that combine visual data and process parameters to recognize early indicators of mass flow anomalies. Addressing this gap is crucial to improving reliability, reducing waste, and ensuring stability in high-speed additive manufacturing.

Figure 1 illustrates a real example of nozzle clogging observed, where powder residue partially obstructs the outlet and compromises flow stability.

Figure 1 – View of the nozzle showing signs of clogging at the exit. Such obstructions can compromise powder flow and deposition stability.



Source: The Author (2025).

1.3 General Objective

To develop and implement a smart sensor system that uses machine learning, camera-based visual data, and numeric process features to enable the real-time prediction of mass flow variations, with the goal of preventing clogging events and ensuring consistent powder feeding during laser deposition processes.

1.4 Specific Objectives

The specific objectives of this work are outlined as follows:

a) To design and integrate a smart sensor system capable of combining camera-based visual data and critical numerical features, such as carrier gas flow and disk speed, in a unified framework, ensuring comprehensive monitoring of powder flow behavior;

b) To develop a machine learning model for analyzing multi-modal data streams to predict clogging events and mass flow variations in real time, with a target MAE < 0,3 in z-score and $R^2 > 0,7$;

c) To validate the system under well-illuminated environmental conditions, using wavelength-specific illumination and optical filters to ensure high-contrast, sharp image acquisition for robust prediction performance;

d) To explore and compare different input configurations (e.g., using side images, coaxial images, and gas flow individually and in combination),

assessing their impact on model performance and computational cost;

e) To perform Grad-CAM-based visual interpretation to understand which regions of the input images contribute most to the model's decisions, ensuring that predictions are based on physically meaningful features;

f) To evaluate the model's ability to interpolate untrained conditions by removing specific disk speed steps from the training set and analyzing how the model generalizes to those regions;

g) To assess the system's overall performance in terms of prediction accuracy, response time, and its ability to enhance process reliability, reduce waste, and optimize material usage;

2 State of the Art

This chapter explores the current state of LMD and EHLA techniques, powder flow monitoring methods, as well as machine learning and neural networks.

2.1 Process

Additive Manufacturing (AM) encompasses a diverse range of technologies that enable the layer-by-layer fabrication of components, as categorized by the ASTM F2792 standard. This classification divides AM into seven primary groups, which include Directed Energy Deposition (DED). (ASTM, 2012)

2.1.1 Directed Energy Deposition

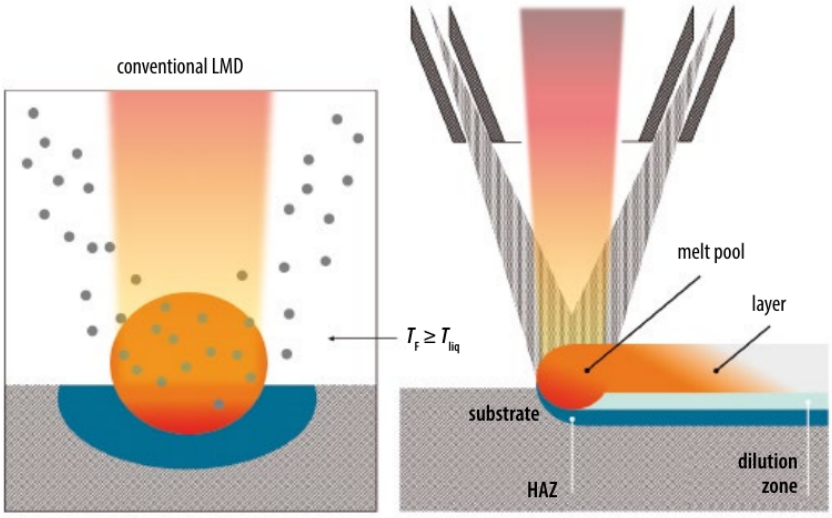
Directed Energy Deposition (DED) refers to a family of additive manufacturing processes that use a focused energy source, such as a laser, plasma arc, or electron beam, to melt and deposit material layer by layer (DASS; MORIDI, 2019). DED stands out for its capacity to deposit metals in various forms, such as powder or wire, directly onto existing surfaces or as new components. Its industrial relevance lies in its versatility, enabling applications such as the repair, reinforcement, and coating of high-value parts. Sectors like aerospace, energy, and manufacturing particularly benefit from DED due to its precision, material efficiency, and ability to handle large-scale parts (ASTM, 2012).

2.1.2 Laser Material Deposition

Among the laser-based DED processes, Laser Material Deposition (LMD) is a well-established technique used for coating and repair applications. LMD employs a laser beam to generate a melt pool on the substrate surface, into which metal powder is injected, creating a metallurgical bond between layers (SCHOPPHOVEN; GASSER; BACKES, 2017). According to Schopphoven, Gasser e Backes (2017), the process typically achieves coating thicknesses between 0.5 and 1 mm and is suitable for repairing components and enhancing wear and corrosion resistance. However, LMD coatings are often too thick for certain applications, and the high heat input results in relatively large heat-affected zones (HAZ), making it less efficient for thin or highly dynamic components (SCHOPPHOVEN; GASSER; BACKES, 2017).

Figure 2 illustrates the conventional LMD process. On the left side of the image, the powder particles are shown melting onto the surface, leading to a higher heat input and resulting in a larger HAZ. On the right side, the image highlights the melt pool, the substrate, and the dilution zone, clearly illustrating the areas where the material interacts with the base surface and where deposition occurs.

Figure 2 – Conventional LMD



Source: Schopphoven, Gasser e Backes (2017).

2.1.3 Extreme High-Speed Laser Material Deposition

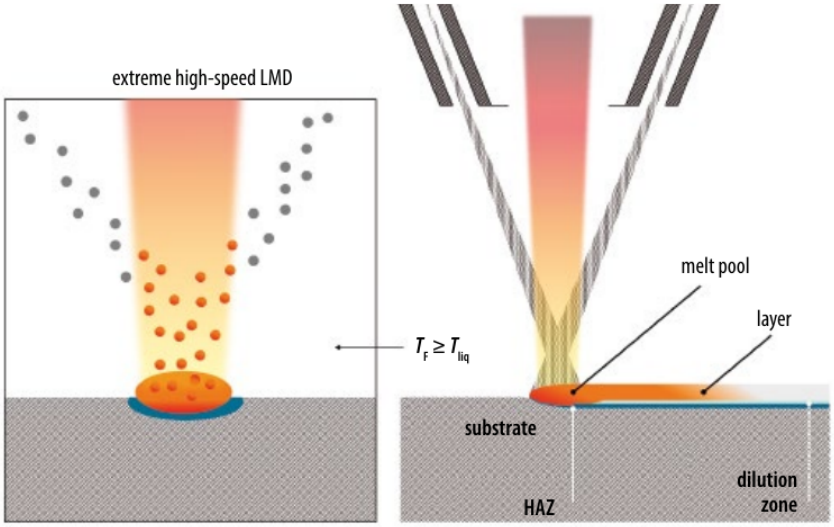
To eliminate the drawbacks of conventional LMD techniques, researchers at Fraunhofer ILT and RWTH Aachen developed the Extreme High-Speed Laser Material Deposition process (EHLA), which offers a more economical solution. EHLA increases the deposition speed from 0.5–2 m/min in LMD to 50–500 m/min, enabling coatings that are 100 to 250 times faster than conventional LMD. Additionally, the heat-affected zone is reduced from 500–1000 µm to just 5–10 µm, minimizing thermal damage to the substrate. (SCHOPPHOVEN; GASSER; BACKES, 2017)

EHLA introduces a significant innovation by melting the powder particles above the melt pool, allowing the production of thin layers between 25 and 250 µm with minimal heat input. These features make EHLA ideal for corrosion and wear protection on large-scale components and heat-sensitive substrates, such as aluminum base alloys or cast iron. Some companies are already using this new process, such as the Dutch firm IHC Vremac Cylinders BV. They use EHLA to coat their hydraulic cylinders, which exhibit a diameter of up to fifty centimeters and a length up to ten meters, for offshore applications. By overcoming the speed and thermal limitations of LMD, EHLA has opened new possibilities for high-precision coatings and sustainable manufacturing. (SCHOPPHOVEN; GASSER; BACKES, 2017)

Figure 3 illustrates the EHLA process. On the left side of the image, the powder particles are seen melting above the component surface. This reduces the laser radiation required on the substrate, since the particles are already molten upon impact. As a result, the process eliminates the need for the additional time increment to remelt particles in the melt pool, allowing for faster processing, reduced energy input, and a smaller heat-affected zone compared to the conventional LMD process. On the right

side, the image highlights the melt pool, the substrate, and the dilution zone, showing the areas where the material interacts with the base surface and where deposition occurs. The contrast between the two processes is evident, with EHLA demonstrating a more controlled deposition and reduced thermal impact on the substrate.

Figure 3 – Extreme High-Speed LMD



Source: Schopphoven, Gasser e Backes (2017).

The choice of LMD and EHLA is motivated by their complementary strengths. LMD provides high precision, making it ideal for intricate repairs and coatings requiring detailed control. Conversely, EHLA's high-speed capabilities allow for the efficient processing of large surfaces or high-volume tasks while maintaining excellent layer quality. Together, these technologies represent a transformative advancement in additive manufacturing, addressing diverse industrial needs, from micro-scale precision to macro-scale production.

2.2 In-situ Monitoring and Data-driven Modeling in Laser Deposition Processes

Reliable monitoring of powder mass flow is essential for ensuring consistent material deposition in DED processes, particularly in techniques like LMD and EHLA. Various methods have been developed to estimate or measure the powder flow rate during operation, though most still present limitations in terms of sampling frequency, robustness, or integration with the process environment.

The lack of advanced, real-time monitoring solutions capable of integrating visual data (e.g., camera footage) with numeric process parameters (e.g., gas flow rate, pressure, and feed rate) exacerbates the problem, limiting the ability to detect and prevent deviations proactively. Bonin *et al.* (2023) emphasize that traditional detection methods are insufficient for identifying subtle anomalies in powder delivery systems, as they often rely on binary fault detection rather than predictive monitoring. This

limitation restricts their ability to address issues preemptively in dynamic manufacturing environments. Similarly, Tang *et al.* (2008) outline several existing approaches for monitoring powder flow rates, each with significant drawbacks. One such method involves weight-loss metering systems, which suffer from low sampling frequencies (e.g., 10 Hz), making them unsuitable for real-time applications. Optoelectronic sensors, though promising in theory, fail under the thermal and spatial constraints near the nozzle. Pressure-based approaches, in turn, are often system-specific and difficult to generalize (TANG *et al.*, 2008).

These limitations underscore the importance of developing data-driven and sensor-integrated solutions capable of interpreting process behavior in real time. Recent advances in machine learning have enabled more robust and adaptive monitoring systems for laser-based deposition processes. In particular, the combination of visual sensing and operational parameters has demonstrated promising results in predicting process outcomes and enhancing monitoring accuracy in laser deposition processes.

Perani *et al.* (2023) propose a deep learning framework that combines coaxial melt pool images with scalar process parameters (e.g., power and speed) to predict the geometry of the deposited track in LMD. Their convolutional architecture integrates visual features and numerical inputs to estimate width, height, and cross-sectional area in real time. This approach demonstrates the effectiveness of multimodal data fusion for on-line geometry prediction, reinforcing the feasibility of using similar models for powder flow estimation.

Similarly, Dong *et al.* (2023) developed a multi-mode convolutional neural network (M-CNN) composed of a ResNet18-based CNN and a fully connected network to predict the geometry of deposited tracks. Their model combines molten pool images, temperature data, and process parameters (e.g., laser power, scanning speed) to estimate width, height, and cross-sectional area. As stated in the paper, "the M-CNN [...] used deposition images, temperature and process parameters as inputs to predict the deposited layer sizes" (DONG *et al.*, 2023, p. 791). Their results confirm that multimodal data fusion significantly improves prediction accuracy, highlighting the relevance of visual and scalar features for real-time process monitoring in LMD.

In another contribution, Zeníšek *et al.* (2022) proposed the use of virtual sensors to approximate the behavior of physical sensors through machine learning models. As explained in their work, “The idea behind this concept is to create a data-based model, using machine learning algorithms, which approximates a certain physically available sensor, by using other data recordings as input features.” This aligns closely with the concept of using coaxial imaging and scalar process parameters to infer latent variables such as powder mass flow, where direct measurements may be challenging or unreliable.

From a system architecture perspective, Banfi *et al.* (2022) proposed the Ground Control framework, a monitoring and control system for LMD that enables synchronized acquisition of multimodal data, including melt pool images, temperature, and machine signals, and stores them in a structured database to support process analysis and machine learning predictive models. In a similar way, this work combines timestamp-synchronized coaxial and lateral images with carrier gas flow data to estimate powder mass flow using a convolutional neural network. With an average prediction time of 7 ms, the system demonstrates the feasibility of real-time monitoring in high-speed laser deposition processes.

While this research focuses on visual and scalar data for continuous process monitoring, other sensor modalities have also been explored for in-situ analysis in laser-based additive manufacturing. For instance, Chen *et al.* (2023) employed acoustic signals in combination with a convolutional neural network to detect specific defects such as cracks and keyhole pores in LDED. Their system integrates signal denoising, feature extraction in time, frequency, and cepstral domains, and MFCC-based CNN classification. Although focused on discrete defect detection rather than continuous monitoring, the study highlights how alternative sensing modalities can be effectively used in real-time environments, reinforcing the versatility of deep learning frameworks for laser deposition processes.

These studies collectively reinforce the potential of real-time, in-situ monitoring in laser deposition processes through the integration of multimodal sensor data and deep learning. By combining visual information with process parameters and leveraging synchronized data streams, such approaches enable more responsive and informed process supervision. This context supports the proposed framework, which leverages synchronized multimodal data and deep learning to estimate powder mass flow with high temporal resolution, advancing the development of intelligent, in-situ monitoring systems for laser-based additive manufacturing.

2.3 Machine Learning and Neural Networks

2.3.1 Overview of Machine Learning

"A machine learning algorithm is an algorithm that is able to learn from data." (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 97). "Machine learning algorithms can be broadly categorized as unsupervised or supervised by what kind of experience they are allowed to have during the learning process." (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 102). These learning paradigms form the foundation of more advanced techniques, including Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs) (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.3.2 Supervised Learning

Supervised learning involves training a model using labeled datasets, where each input corresponds to a known output. The goal is to learn a mapping function that minimizes errors in prediction for unseen data. This technique is particularly effective for tasks such as image classification, speech recognition, and natural language processing. Models like linear regression, support vector machines, and neural networks are commonly employed in supervised learning. (GOODFELLOW; BENGIO; COURVILLE, 2016; BISHOP, 2006)

In the context of powder-based additive manufacturing, supervised learning can be applied to predict operational variables that are directly linked to powder flow behavior. In this study, a regression model was developed to estimate the disk speed, a parameter that correlates strongly with mass flow, based on high-speed process images and the carrier gas flow rate. By learning from historical data, the model captures the underlying dynamics of the deposition system. Deviations between predicted and actual disk speed values may reveal indirect evidence of process anomalies such as partial clogging, powder leakage, or irregular feeding conditions. For example, if the system is expected to deliver 15 g/min and the predicted value suddenly drops to 12 g/min under otherwise stable conditions, this may suggest an obstruction or system leakage. Rather than relying on binary labels (e.g., “normal” or “anomalous”), this approach provides a more nuanced, data-driven mechanism for real-time process monitoring and early anomaly detection.

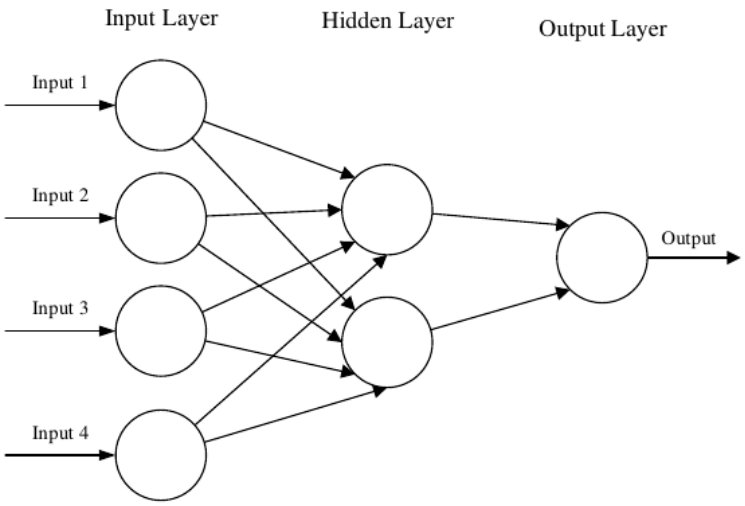
2.3.3 Neural Networks

Artificial Neural Networks (ANNs) are computational systems designed to mimic the functioning of biological nervous systems, such as the human brain. These networks are composed of numerous interconnected computational units, commonly referred to as neurons, which collaborate in a distributed manner to process input data and optimize the final output. This process of optimization, known as learning, is central to how ANNs operate. (O’SHEA; NASH, 2015)

The basic structure of an ANN, illustrated in Figure 4, consists of three main types of layers: input, hidden, and output. The input layer is the first stage of the network, where data is introduced, each node corresponds to a specific input feature (e.g., Input 1, Input 2, etc.), and in this case, the network accepts four features. The hidden layer processes these inputs using weighted connections and non-linear activation functions: each node combines the signals it receives, applies a transformation, and adjusts its internal parameters to learn patterns or relationships in the data. Arrows between nodes indicate the flow of information through the network, with each connection having an associated weight that determines the strength of influence between nodes. These

weights are refined iteratively during training through gradient-based optimization. The output layer produces the network's final result, represented here by a single output node, indicating a scalar output for tasks such as regression. When multiple hidden layers are stacked between the input and output layers, the model becomes a deep neural network and is said to engage in deep learning. (O'SHEA; NASH, 2015)

Figure 4 – Basic Artificial Neural Network (ANN) architecture with three distinct layers



Source: O'Shea e Nash (2015).

2.3.4 Fundamentals of Learning in Neural Networks

The learning process in neural networks is grounded in core mathematical principles that allow the model to adjust its internal parameters in response to data. One of the foundational concepts is the perceptron, introduced by Rosenblatt (1958) and further elaborated in modern deep learning literature by Goodfellow, Bengio e Courville (2016). It models a single neuron that takes multiple inputs, applies weights, adds a bias, and processes the result through an activation function. The output of a perceptron is determined by the weighted sum of inputs and the applied nonlinear activation, such as the sigmoid, tanh, or ReLU (Rectified Linear Unit) functions. These nonlinearities are essential for enabling networks to approximate complex, non-linear relationships that linear models cannot capture. (ROSENBLATT, 1958; GOODFELLOW; BENGIO; COURVILLE, 2016)

Once a network is defined, it must be trained to perform a specific task, such as regression or classification. This is accomplished through supervised learning, where the model iteratively adjusts its parameters based on the discrepancy between the predicted outputs and the ground truth labels. This discrepancy is quantified by a loss function, such as mean squared error (MSE) for regression tasks or cross-entropy for classification. (GOODFELLOW; BENGIO; COURVILLE, 2016)

The mechanism by which neural networks learn is known as backpropagation, an algorithm that applies the chain rule to compute how each parameter in the network influences the loss function. This is done by propagating the loss gradients backward through the layers, allowing the model to update each parameter based on its contribution to the total error. These gradients are then used in conjunction with gradient descent, an optimization method that adjusts the parameters in the direction that minimizes the loss function. (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 199-208, p. 274-276) A commonly used didactic form of the weight update rule in gradient descent is:

$$w_{t+1} = w_t - \eta \cdot \frac{\partial L}{\partial w} \tag{1}$$

Where η is the learning rate, a hyperparameter that controls the step size of each update, and $\frac{\partial L}{\partial w}$ is the gradient of the loss L with respect to the parameter w. While this is a simplified form, it reflects the core idea of the general gradient-based update rule presented by Goodfellow, Bengio e Courville (2016, p. 276).

$$\theta \leftarrow \theta - \eta \nabla_\theta J(\theta) \tag{2}$$

Choosing an appropriate learning rate is critical for the effectiveness of training. As Goodfellow, Bengio e Courville (2016, p. 278) explain, if the learning rate is too low, learning proceeds slowly or may become stuck with a high-cost value. On the other hand, overly large learning rates can lead to unstable updates and divergence during training, particularly in deep architectures. (GOODFELLOW; BENGIO; COURVILLE, 2016)

Several enhancements to basic gradient descent have been developed to improve performance and convergence, including momentum, RMSProp, and Adam optimizers. These methods dynamically adjust the learning step or incorporate memory of previous updates to better navigate complex loss surfaces and escape local minima. (GOODFELLOW; BENGIO; COURVILLE, 2016)

Together, these mechanisms form the foundation of learning in neural networks, enabling models to generalize from training data and capture complex dependencies, which are essential for tasks like mass flow estimation and powder feed monitoring in advanced manufacturing processes.

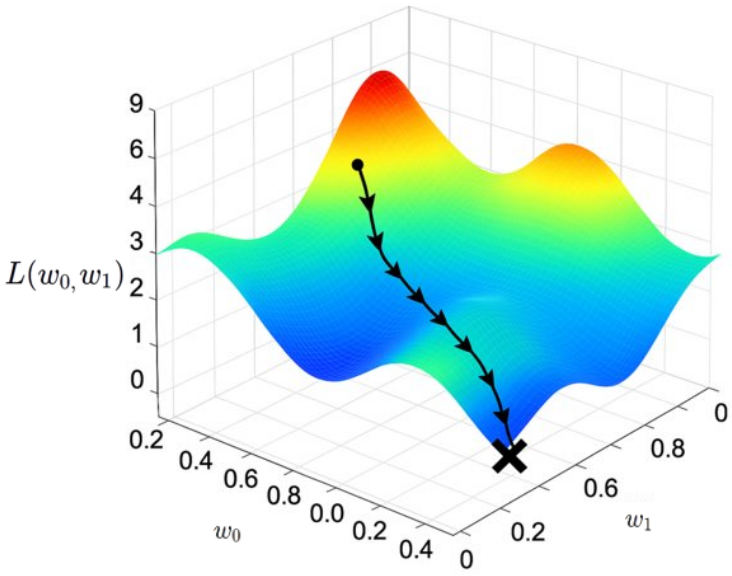
2.3.4.1 Gradient Descent and Backpropagation

Once the loss function has been defined for a neural network, the process of minimizing it relies primarily on two complementary mechanisms: gradient descent and backpropagation.

The gradient descent algorithm is an optimization method that minimizes a loss function $L(\theta)$, where θ represents the parameters of the model. The algorithm computes the gradient $\nabla_\theta L$, which points in the direction of greatest increase in the loss and then updates the parameters in the opposite direction. (GOODFELLOW; BENGIO; COURVILLE, 2016)

Figure 5 illustrates the conceptual idea of gradient descent descending a loss landscape toward a local minimum.

Figure 5 – Gradient descent descending a loss landscape toward a local minimum



Source: The Author.

Backpropagation complements this process by efficiently computing the partial derivatives of the loss function with respect to all weights in the network. It does so by applying the chain rule of calculus recursively, starting from the output layer and propagating backward through the hidden layers. This makes it computationally feasible to train deep networks with millions of parameters. (GOODFELLOW; BENGIO; COURVILLE, 2016)

In each training iteration, the following steps occur:

1. *Forward pass*: the input propagates through the network, generating a prediction.
2. *Loss computation*: the output is compared to the ground truth using a loss function.
3. *Backward pass*: the gradients of the loss with respect to each parameter are computed via backpropagation.
4. *Parameter update*: weights and biases are updated using gradient descent.

Different variants of gradient descent exist to enhance convergence and generalization. These include:

- *Stochastic Gradient Descent (SGD)*: updates using one sample at a time.
- *Mini-batch Gradient Descent*: updates using small batches, balancing speed and stability.
- *Adaptive methods* (e.g., Adam): adjust learning rates for each parameter during training, often leading to faster convergence and better generalization in practice (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 311-313).

While gradient descent may converge to local minima or saddle points, empirical evidence suggests that these often yield acceptable generalization performance in deep networks (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 300-304).

2.3.5 Convolutional Neural Networks (CNNs)

CNNs share similarities with traditional ANNs in that they consist of neurons that self-optimize through learning. Each neuron processes inputs and applies operations like scalar products and non-linear functions, which are the foundation of countless ANN architectures (Haykin, 2008). From the raw input image vectors to the final output (e.g., class scores), CNNs operate as end-to-end systems that optimize weights and minimize associated loss functions (GOODFELLOW; BENGIO; COURVILLE, 2016).

The distinguishing feature of CNNs lies in their suitability for pattern recognition within images. Krizhevsky, Sutskever e Hinton (2012) describe their network as comprising “five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers,” featuring “60 million parameters and 650,000 neurons.” Unlike traditional ANNs, the architecture of CNNs, with its specialized convolutional layers, inherently encodes image-specific features. This design significantly reduces the number of parameters; indeed, the authors note that “compared to standard feedforward neural networks with similarly-sized layers, CNNs have much fewer connections and parameters and so they are easier to train,” making them highly efficient and well-suited for large-scale image classification tasks. Furthermore, CNNs, as highlighted by LeCun, Bengio e Hinton (2015), have brought about breakthroughs in processing images, video, speech, and audio, cementing their role as a cornerstone in modern computer vision tasks.

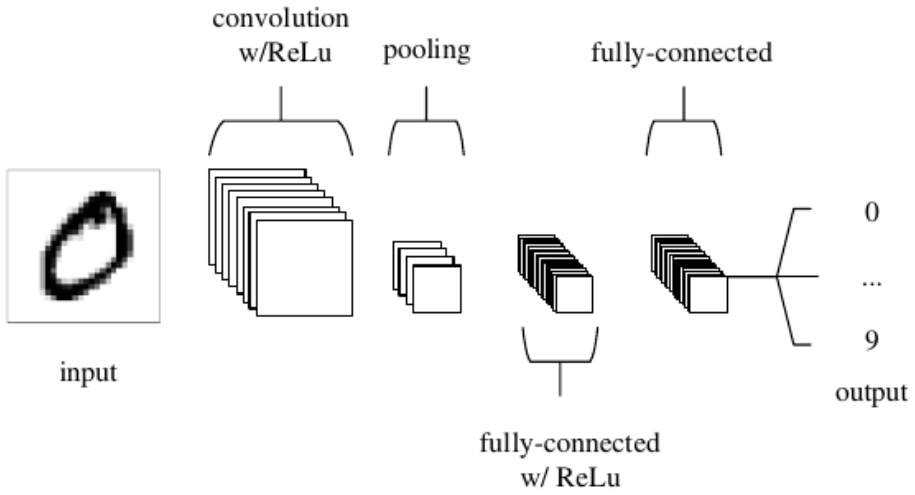
CNN architecture

A typical CNN alternates between convolutional and pooling layers before feeding the data into one or more fully connected layers. This hierarchical structure progressively abstracts features from low-level edges to high-level patterns, as depicted in simplified diagrams (O’SHEA; NASH, 2015, p. 5).

Figure 6 illustrates a simple CNN architecture, consisting of three main types of layers, convolutional layers, pooling layers, and fully connected layers, that work

together to process image data. These layers are typically stacked to form the CNN architecture.

Figure 6 – A simple CNN architecture, comprised of just five layers



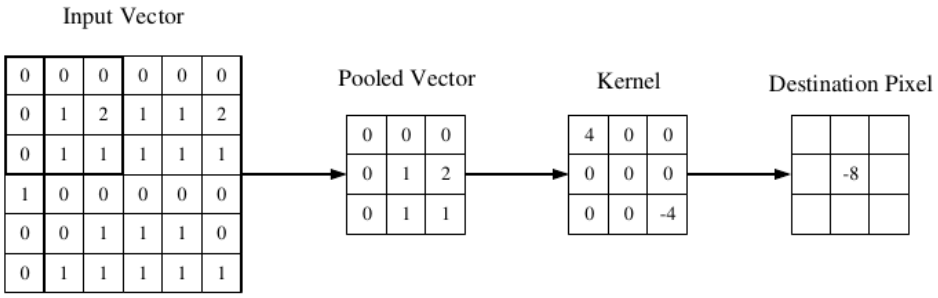
Source: O'Shea e Nash (2015).

Convolutional layers

The convolutional layer is the cornerstone of CNNs. It applies learnable filters (kernels) to local regions of the input, generating activation maps that highlight specific features such as edges, textures, or patterns. The local connectivity of convolutional layers enables a reduction in the number of trainable parameters compared to fully connected layers, making CNNs computationally efficient (LECUN *et al.*, 1998; O'SHEA; NASH, 2015). The process of convolution allows the network to detect spatial hierarchies, which are critical for image classification tasks.

In Figure 7, a visual representation of a convolutional layer can be seen. The center element of the kernel is placed over the input vector, of which is then calculated and replaced with a weighted sum of itself and any nearby pixels.

Figure 7 – A visual representation of a convolutional layer



Source: O'Shea e Nash (2015).

Pooling Layers

Pooling layers reduce the spatial dimensions of activation maps, preserving the most important features while lowering computational demands. Max-pooling, the most common approach, selects the maximum value within a region to downsample the data. By reducing the resolution, pooling layers also help prevent overfitting, a common issue in deep networks. (LECUN *et al.*, 1998; O'SHEA; NASH, 2015)

Fully Connected Layers

Fully connected layers aggregate features extracted from previous layers to perform high-level reasoning, such as classification. These layers connect every neuron from the preceding layer to produce output probabilities for different classes. Functions like ReLU (Rectified Linear Unit) or softmax are typically applied to improve performance and ensure the outputs are interpretable as probabilities. (O'SHEA; NASH, 2015)

2.4 Gradient-weighted Class Activation Mapping (Grad-CAM)

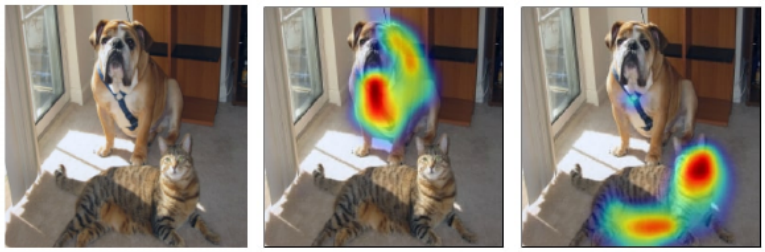
Gradient-weighted Class Activation Mapping (Grad-CAM) is a widely used technique for visualizing and interpreting the decision-making process of CNNs. Proposed by SELVARAJU (2017), Grad-CAM generates visual explanations by computing the gradient of the output class score with respect to the feature maps of a convolutional layer. These gradients are used to produce a coarse localization map, highlighting the regions in the input image that are most influential for the model's prediction (SELVARAJU, 2017).

The Grad-CAM process involves three main steps: (1) compute the gradient of the class score with respect to the feature maps, (2) apply global average pooling to generate importance weights for each feature map, and (3) produce a weighted sum of the feature maps followed by a ReLU activation to obtain the final heatmap.

This technique is particularly useful for debugging CNNs, verifying whether the model focuses on meaningful regions, and increasing transparency in decision-making. In manufacturing applications, Grad-CAM can help assess whether the model correctly focuses on the melt pool and powder stream, or is distracted by noise or background elements.

To better understand how Grad-CAM highlights class-discriminative regions, Figure 8 shows an illustrative example adapted from SELVARAJU (2017). The original image contains both a cat and a dog. The corresponding Grad-CAM visualizations for each class demonstrate how the network attends to different regions depending on the predicted label.

Figure 8 – Grad-CAM class-discriminative localization. (a) Original image. (b) Grad-CAM heatmap for the class "dog". (c) Grad-CAM heatmap for the class "cat".



Source: Adapted from SELVARAJU (2017).

Extensions such as Grad-CAM++, proposed by CHATTOPADHYAY (2018), and Guided Grad-CAM combine this approach with other gradient-based methods to provide finer-grained visualizations and improved localization, particularly in multi-object scenes.

2.5 Ludwig AI

Ludwig AI is an open-source, declarative machine learning framework designed to simplify workflows by enabling users to develop and deploy deep learning models without extensive coding. The framework allows users to define machine learning pipelines using YAML configuration files, making it accessible even to those with limited programming expertise. Ludwig supports tasks such as natural language processing, computer vision, and multimodal learning, bridging simplicity and advanced functionality. (LUDWIG AI, 2024b; LUDWIG AI, 2024a)

A relevant application of multimodal deep learning in industrial contexts is exemplified by the work of Bauer *et al.* (2023), who combined thermal images and camera images to monitor the quality of 3D printing processes in additive manufacturing. Their approach involved fusing heterogeneous data sources, thermal sensor data and RGB visual data, using a deep learning model to detect defects and ensure print quality. Although frameworks like Ludwig AI are not explicitly cited, their declarative and multimodal architecture would be highly suited for replicating such experimental setups. This study demonstrates the value of multimodal approaches in complex industrial environments, emphasizing the need for robust data integration strategies.

Similarly, Ludwig's multimodal learning framework has significant potential in powder-based additive manufacturing. The ability to combine image data from cameras with numeric features aligns closely with Ludwig's design principles. By leveraging its declarative approach and flexibility, it is possible to create robust models capable of predicting critical phenomena such as clogging events or mass flow variations, showcasing its adaptability to industrial contexts.

3 METHODOLOGY

This chapter describes the methodology adopted for the development and validation of the smart sensor system. Section 3.1 provides an overview of the general approach. Section 3.2 presents the experimental setup. Section 3.3 describes the data acquisition and preparation steps, including labeling and preprocessing. Finally, Section 3.4 outlines the model training and evaluation process, and discusses the potential for real-time integration with the EHLA process.

3.1 Methodology Overview

This research aims to develop and implement an intelligent sensor system capable of real-time prediction of mass flow variations during the EHLA process. The goal is to integrate machine learning, high-speed visual data, and numerical process parameters to anticipate clogging events and ensure consistency in powder-based deposition.

The study is classified as applied and experimental research, as it involves practical testing and applies theoretical concepts, specifically CNNs, to an industrial process. Data collection was carried out through experiments involving the acquisition of coaxial and side images using high-speed cameras (200 and 520 fps), combined with powerful LED illumination and optical filters. This setup was designed to accurately capture the process features, such as the powder cone and the melt pool.

The tools and technologies employed include the Ludwig AI framework, built on top of PyTorch, which simplifies the development and training of machine learning models by relying on configuration files and structured datasets. Image acquisition and processing were performed using Python on a workstation equipped with a Linux Ubuntu system, Intel Xeon processor, RTX 3070 GPU (8GB VRAM), and 64 GB RAM. Additional equipment included a high-intensity 634 nm LED light source, 650 nm bandpass optical filters, Basler industrial cameras, a Fanuc 31i-B5 CNC controller, and a Metco Twin 150 powder feeder.

Data analysis relies on image processing and deep learning techniques. CNNs are trained to recognize visual patterns associated with different mass flow conditions. The choice of methods is based on the ability of high-speed cameras to capture fine details of the process, as well as the efficiency of Ludwig AI in facilitating the development pipeline without extensive manual coding.

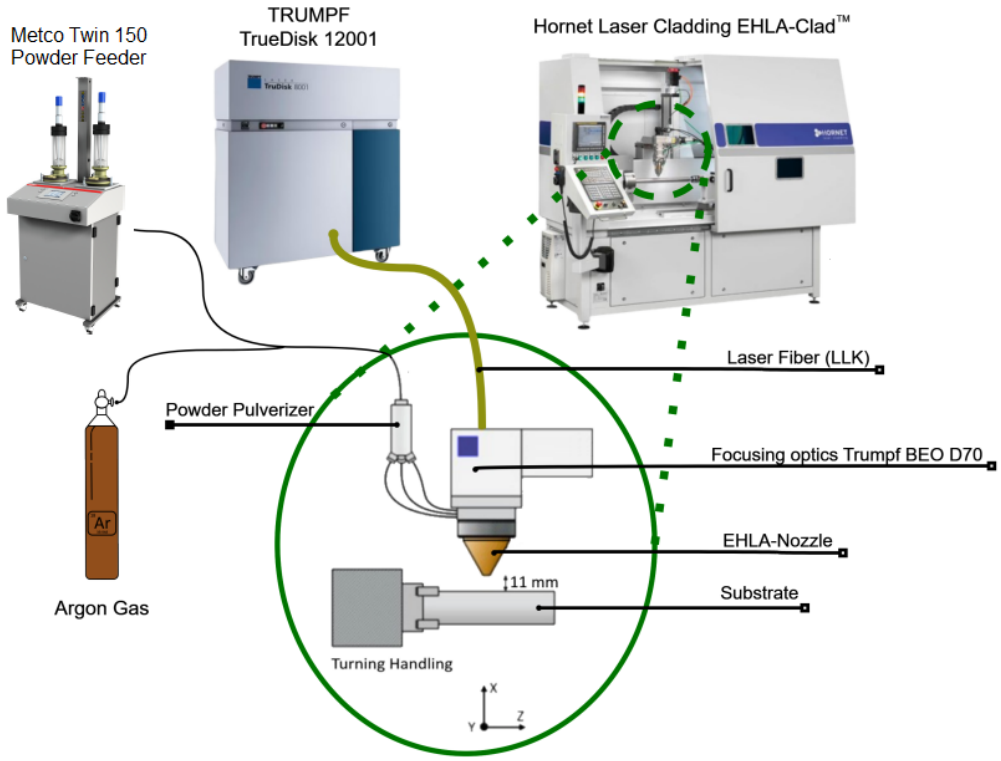
It is important to note that this research serves as a proof of concept. The models were trained using data acquired under specific experimental conditions. Variations in camera positioning, lighting, and environmental factors can significantly affect the quality and generalizability of the results. Further experiments under diverse con-

ditions would be required to improve the robustness and applicability of the proposed system.

3.2 Experimental Environment

The EHLA setup (Figure 9) consists of several key components essential for the process. It includes a CNC motion system, a laser source, an optical system, a powder feeder, and a deposition nozzle. High-speed side and coaxial cameras were integrated into the system to enable synchronized image acquisition during the experiments.

Figure 9 – EHLA Setup



Source: Adapted from OERLIKON (2025), TRUMPF (2025), HORNET (2025), and FRAUNHOFER ILT (2024).

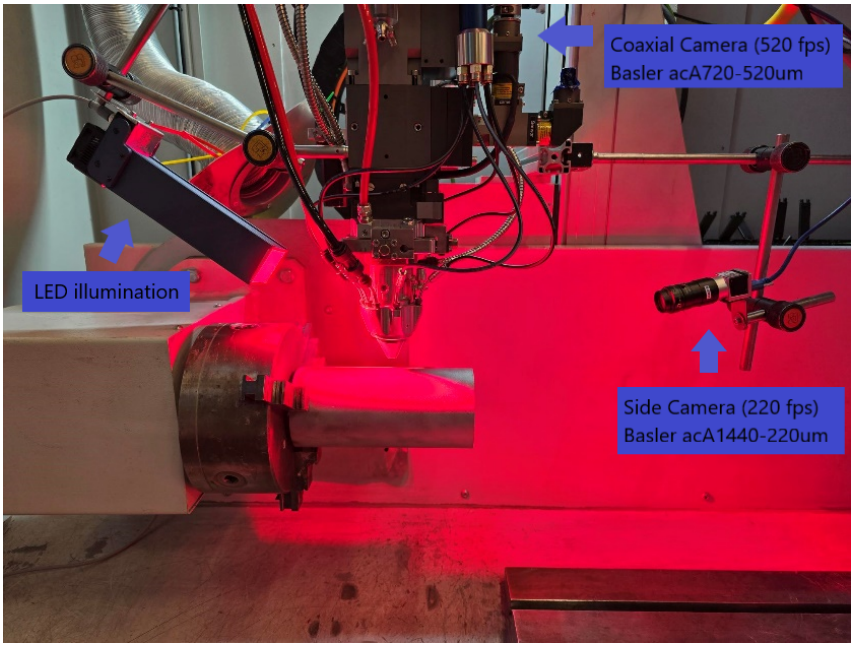
The CNC machine (Hornet Laser Cladding EHLA-Clad) was equipped with a Fanuc Series 30i Model B5 controller and operated with a 4-axis configuration, enabling precise movement and substrate rotation. The powder was delivered using the Oerlikon Twin 150 feeder, which controls mass flow through a dosing disk mechanism. A TRUMPF 12 kW fiber laser was used to generate the beam, which was guided to the deposition area via an optical fiber and focused using a galvo mirror and F-theta lens. Argon gas was used as both the carrier and shielding gas, ensuring proper powder transport and melt pool protection.

High-speed imaging was performed using two Basler industrial cameras, the acA1440-220um and the acA720-520um. The acA1440-220um offers a resolution of 1448 × 1088 pixels (1.6 MP) at up to 227 FPS, while the acA720-520um provides 728 × 544 pixels (0.4 MP) at up to 525 FPS. Both models use Sony CMOS sensors and connect via USB 3.0. These sensors share a peak spectral sensitivity near 600 nm (see Figure 44 in Annex 5.2), which aligns with the 634 nm illumination and the 630–670 nm filter bandwidth used in the setup, ensuring high-quality signal capture in the red region of the visible spectrum (Figure 45, Annex 5.2). To enhance imaging contrast and suppress background noise, an optical bandpass filter (FBH650-40) with a central wavelength (CWL) of 650 nm and a full width at half maximum (FWHM) of 40 nm was used. Illumination was provided by a high-intensity 634 nm LED system, selected for its spectral alignment with both the filter and camera sensitivity, enabling robust image acquisition under thermal process conditions.

3.2.1 Sensor Positioning

To support image acquisition and ensure consistent visual monitoring of the EHLA process, the relative positioning of sensors and illumination was carefully defined. Figure 10 shows the actual machine setup, including the final layout of the sensor and illumination system used during data collection.

Figure 10 – Machine setup used for data acquisition



Source: The Author.

In this configuration, the coaxial camera is aligned with the laser beam path, positioned directly above the melt pool to capture the interaction zone from a top-down perspective. The side-view camera is mounted laterally, angled to capture the powder

jet and its interaction with the laser beam. The LED illumination system is positioned diagonally, directing light along the axis of the coaxial camera. This orientation enhances backscattered light capture in the coaxial view while also projecting light toward the powder-laser interaction zone from the side, thereby improving contrast and powder visibility in both camera perspectives, even under low powder flow conditions.

This arrangement was designed to balance the field of view, contrast, and sensitivity across both camera perspectives, while minimizing interference from ambient light and reflections.

3.3 Data Acquisition and Preparation

This section describes the experimental setup, materials, acquisition procedures, and dataset construction pipeline used to train and evaluate the CNN model.

3.3.1 Experimental Setup and Materials

The experiments were conducted using the previously described EHLA system. The deposited material was a 3.50 alloy powder with a particle size distribution of 20–45 µm, delivered through a dosing groove of 11 mm × 0.6 mm on the rotating disk. Depositions were performed on 25CrMo4 steel tubes with an outer diameter of 95.5 mm. The imaging and illumination components followed the configuration detailed in Section 3.2.

3.3.2 Powder Mass Flow Measurement

To obtain a reference for the actual powder mass flow rate, a series of controlled collection tests were conducted. A dedicated collection bottle was positioned below the powder outlet to capture the material dispensed at different disk speed settings. For each test, the powder was collected over a fixed time interval using a stopwatch, and the total mass was measured using a precision scale (Kern PNS 3000-2). The resulting mass flow rate (in g/min), shown in Table 1, was calculated by dividing the measured mass by the elapsed time. This procedure was repeated for multiple disk speed values.

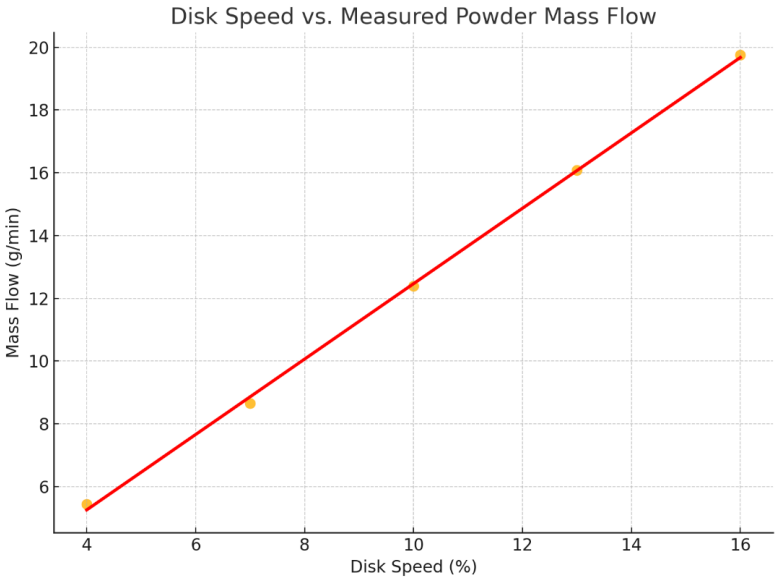
Table 1 – Powder Mass Flow Measurements

Disk speed	Scale value
4%	5.45 g
7%	8.65 g
10%	12.40 g
13%	16.08 g
16%	19.75 g

Source: The Author.

A reference curve relating disk speed (%) to mass flow (g/min) was then plotted to visualize this relationship, as shown in Figure 11. Although minor deviations were observed, the relationship was approximately linear within the operational range of interest. This curve was used for validation and interpretation of the disk speed predictions made by the CNN model in later sections.

Figure 11 – Reference curve: disk speed vs. measured powder mass flow



Source: The Author.

3.3.3 Experimental Protocol and Parameter Selection

The experimental setup underwent iterative refinement throughout the data acquisition process to optimize image quality, powder visibility, and parameter coverage. Initial experiments were designed to explore a broad range of operating conditions and assess the system's sensitivity to key variables such as disk speed and carrier gas flow. However, early results revealed limitations in model performance when trained on overly broad or visually inconsistent datasets. Consequently, both the acquisition parameters and the optical setup were progressively adjusted to improve powder stream visualization and ensure data consistency.

The final dataset used for training and evaluation (Week 2) was acquired under this optimized configuration. Tables 3 to 6, in the annex section, summarize the key differences between the first and second experimental campaigns.

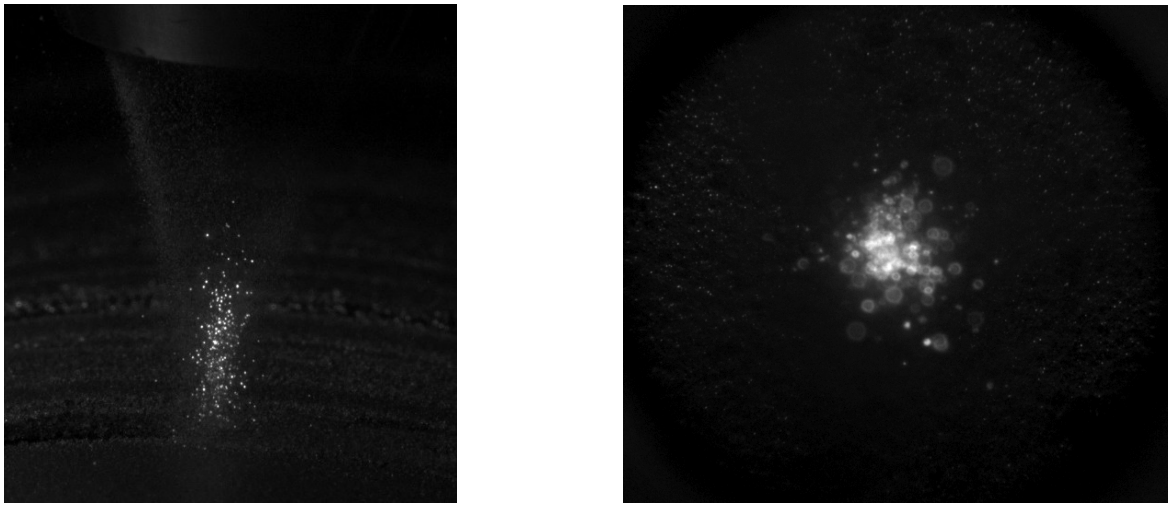
3.3.3.1 Week 1 – Exploratory Phase

In the first campaign, disk speeds ranged from 0% to 100%, with carrier gas flow rates between 4 and 16 normal liters per minute (NLPM). These conditions resulted in powder mass flow rates of approximately 0–120 g/min. Claddings were performed

with a width of 5 mm and a travel speed of 0.5 mm/rev, using a fixed laser power of 3000 W and a process speed of 80 m/min.

Due to the high powder concentration, both cameras captured rich visual data. The side-view camera provided a clear view of the full injection cone (Figure 12a), while the coaxial view, aligned along the laser axis, focused on the interaction zone but often suffered from overexposure and glare due to intense backscatter (Figure 12b).

Figure 12 – Week 1 - with illumination and filters: (a) Side View and (b) Coaxial View

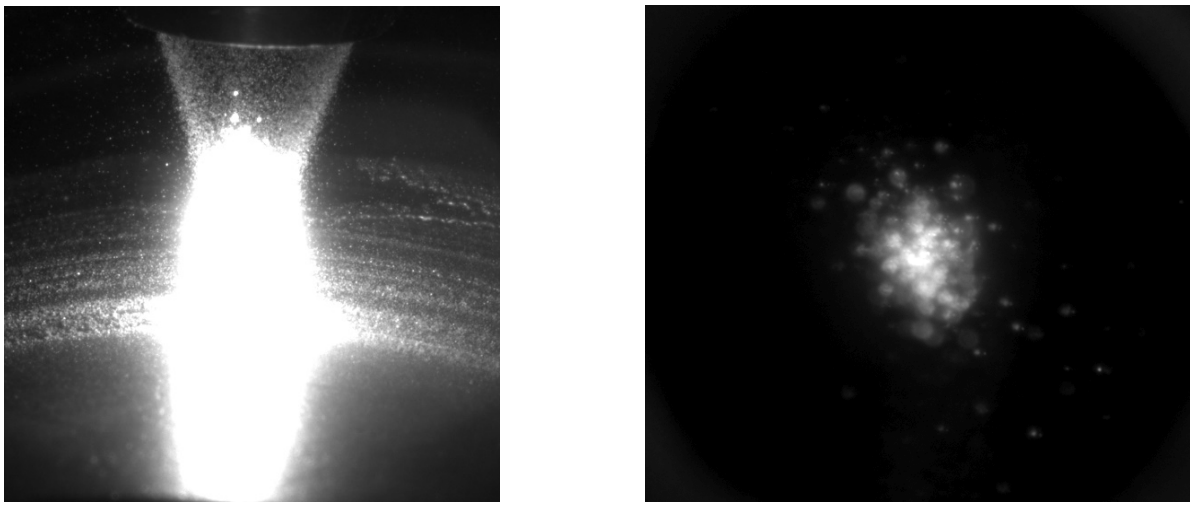


(a) Side View (b) Coaxial View

Source: The Author.

Several lighting setups were tested, including configurations without filters or external illumination, given the high powder visibility at this stage. However, these settings led to overexposed images. Figure 13a presents the side view under such conditions, while Figure 13b shows the corresponding coaxial view.

Figure 13 – Week 1 - without illumination and filters: (a) Side View and (b) Coaxial View



(a) Side View (b) Coaxial View

Source: The Author.

Although useful for early-stage observations, the configurations without illumination and filters led to poor image quality. Moreover, the process was unstable in several claddings, frequently resulting in lack of fusion or irregular melt tracks (Figure 14). Combined with the sparse parameter variation, these issues limited the potential for effective learning in a regression setting.

Figure 14 – Tube Week 1 - Lack of fusion

Source: The Author.

A total of 64 claddings were performed, generating approximately 1,600 coaxial images and 700 side-view images per cladding:

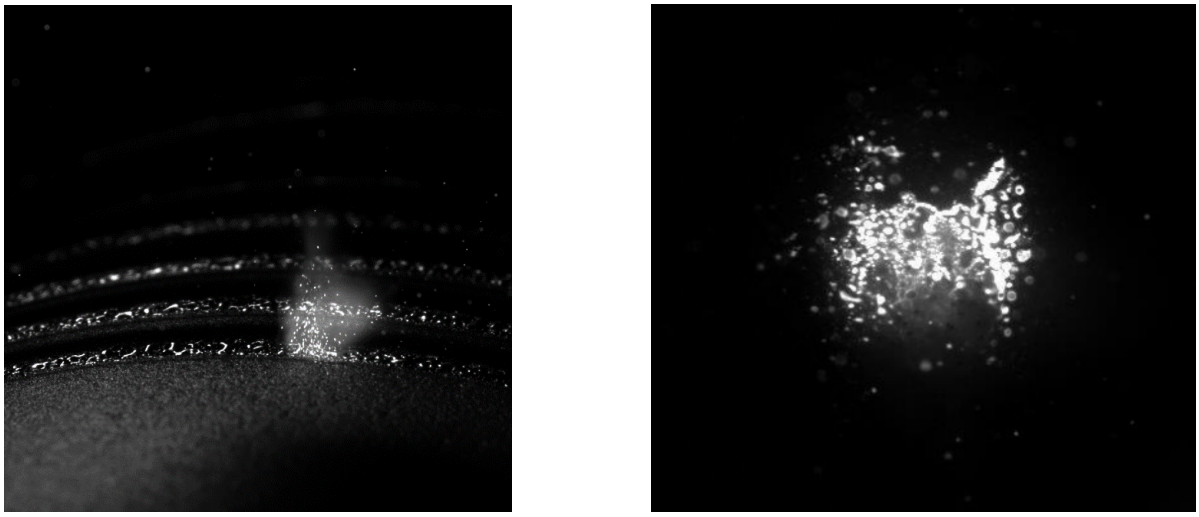
- Coaxial: 102,400 images
- Side-View: 44,800 images

3.3.3.2 Week 2 – Optimized Setup (Final Dataset)

In the second week, the parameter range was refined to 4–16% disk speed, with 3% increments, corresponding to a powder mass flow range of approximately 5–20 g/min. Cladding width was reduced to 3 mm, and travel speed was set to 0.1 mm/rev, resulting in longer deposition times and a higher density of image frames per experiment.

While this range better represented practical operation conditions, with the reduced powder flow, the upper portion of the injection cone became less visible in the side view (Figure 15a), but the interaction zone between the laser and powder particles remained clearly visible, especially in the lower region of the cone, where fine details could still be observed.

Figure 15 – Week 2: (a) Side View and (b) Coaxial View

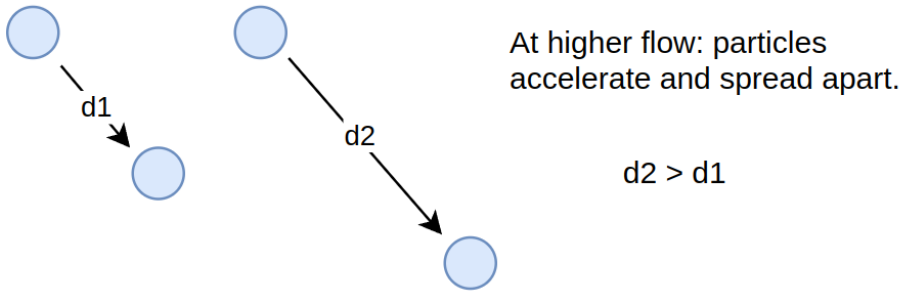


(a) Side View (b) Coaxial View

Source: The Author.

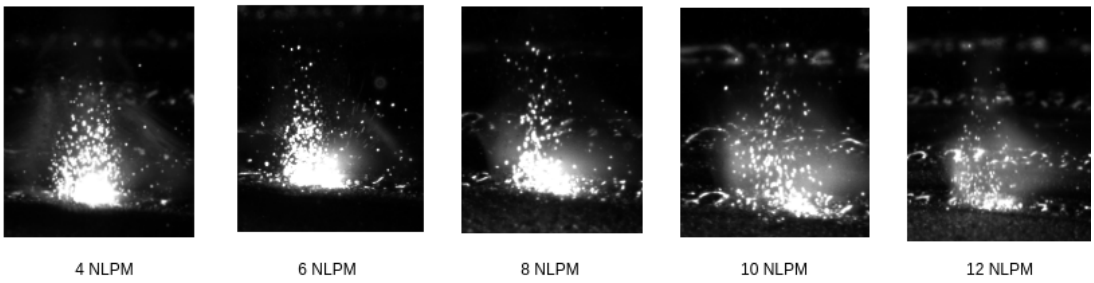
Carrier gas flow was varied across 4, 6, 8, 10 and 12 NLPM, introducing different powder jet behaviors, a key factor considered in model design. Specifically, higher gas flow rates led to greater dispersion and acceleration of powder particles (as illustrated in Figure 16), resulting in visually distinct jet geometries even at similar mass flow rates (Figure 17). For this reason, carrier gas flow was included as a numeric input to the model alongside image data, enabling the CNN to disambiguate flow conditions that produce similar mass flow but different visual patterns.

Figure 16 – Particle Dispersion Increases with Carrier Gas Flow - Illustrative representation showing how particle dispersion varies with carrier gas flow. Distances d_1 and d_2 represent the spacing between particles under different flow conditions, with $d_2 > d_1$.



Source: The Author.

Figure 17 – Impact of Carrier Gas Flow on Powder Jet Behavior (10% Disk Speed)

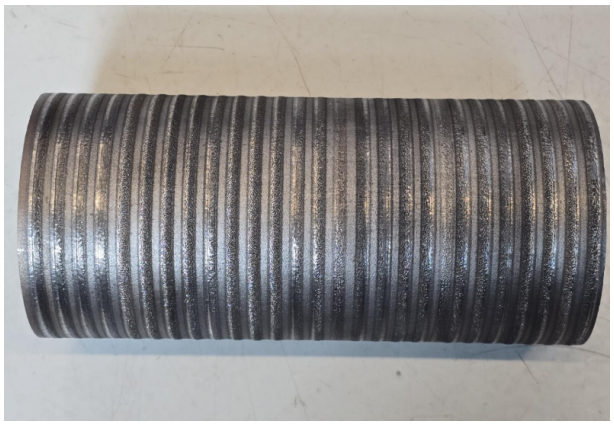


Source: The Author.

A total of 35 claddings were performed under these conditions, generating approximately 3,400 coaxial images and 1,700 side-view images per cladding:

- Coaxial: 119,000 images
- Side-View: 59,500 images

Figure 18 – Tube produced during Week 2 using optimized parameters. Unlike earlier trials in Week 1, this deposition shows improved layer fusion and consistent material buildup.



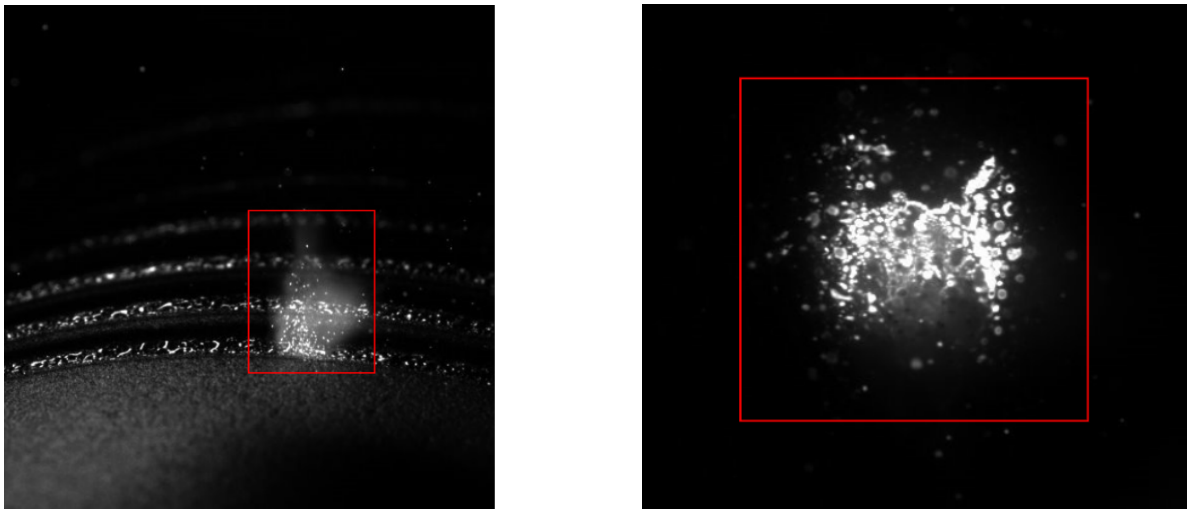
Source: The Author.

3.3.4 Dataset Construction and Labeling

The final dataset was constructed using synchronized image pairs (coaxial and side view), each associated with a disk speed value and carrier gas flow input. The process was automated using a custom Python script (csv_filler.py), which:

1. Parsed timestamps from .tiff image metadata
2. Matched coaxial and side-view frames based on minimal timestamp difference
3. Applied fixed ROIs for each view, followed by cropping and resizing to 240×240 pixels
4. Wrote image paths and parameters into a unified CSV file

Figure 19 – Week 2: (a) Side View ROI and (b) Coaxial View ROI



(a) Side View ROI

(b) Coaxial View ROI

Source: The Author.

Since each cladding was executed with fixed disk speed and carrier gas flow values, the label (disk speed) was extracted directly from the G-code associated with each deposition. No interpolation was required. Each row in the dataset represented a synchronized image pair with its corresponding process parameters: `side_path` and `coaxial_path` (image inputs), `carrier_gas` (numerical input), and `disk_speed` (regression target).

A total of approximately 30,000 images (15,000 synchronized pairs) were used, selected uniformly from the Week 2 dataset to maintain distributional balance across parameter values. Specifically, 3,000 pairs were used for each step (4%, 7%, 10%, 13%, and 16%).

3.4 Model Development and Integration

The core objective of this work was to develop a deep learning model capable of estimating powder mass flow during the EHLA process based on visual input and process parameters. To achieve this, a CNN was designed and implemented using the Ludwig AI framework, which allows for multimodal model configuration via declarative YAML files. The model processes synchronized side-view and coaxial images, along with the corresponding carrier gas flow value, to predict disk speed as a continuous output in a supervised regression setting.

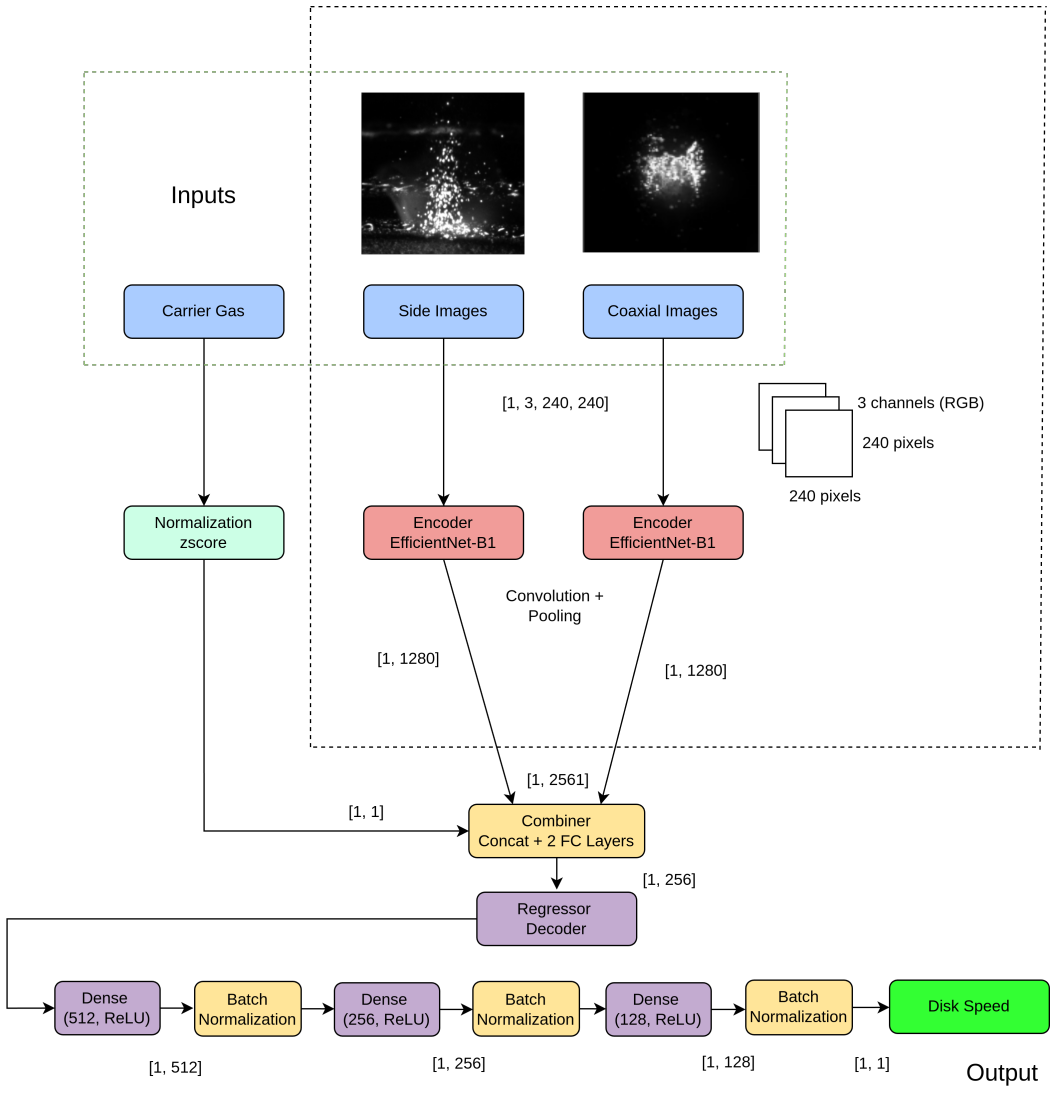
This section outlines the design and structure of the model architecture, its implementation details, training configuration, evaluation metrics, and the steps taken to enable potential real-time integration with the EHLA system.

3.4.1 Architecture Design

Building on the general approach described above, the final CNN architecture was designed to handle heterogeneous input types, two high-speed grayscale image streams and one numerical value, by combining separate encoders with a unified regression head. Each input image (240×240 pixels, single-channel) was processed by a dedicated convolutional encoder. The carrier gas flow, treated as a scalar numerical input, was fed directly into the model without convolutional processing. The resulting feature vectors were concatenated and passed through a sequence of fully connected layers to produce a single numerical output corresponding to the predicted disk speed.

Figure 20 illustrates the complete model architecture, including the three input modalities, their respective encoding paths, the feature concatenation step, and the dense regression output block.

Figure 20 – Model architecture



Source: The Author.

Input Features

The model receives the following inputs:

- Side-view image (`side_path`): grayscale, 240×240 pixels
- Coaxial image (`coaxial_path`): grayscale, 240×240 pixels
- Carrier gas flow (`carrier_gas`): scalar float input (NLPM)

Each image input was processed using a separate EfficientNet-B1 encoder pretrained on ImageNet and fine-tuned during training. A dropout rate of 0.4 was applied, and both encoders were set as trainable to enable domain adaptation to EHLA-specific visual characteristics. The carrier gas input was passed directly to the concatenation layer as a numerical feature.

Feature Combination and Output

The outputs of both image encoders and the numerical input were concatenated using a concat combiner and passed through a series of dense layers. The architecture of the regression head consisted of:

- Dense (512 units) $\rightarrow$ ReLU $\rightarrow$ Dropout (0.4) $\rightarrow$ L2 regularization
- Batch normalization
- Dense (256 units) $\rightarrow$ ReLU $\rightarrow$ Dropout (0.4) $\rightarrow$ L2 regularization
- Batch normalization
- Dense (128 units) $\rightarrow$ ReLU $\rightarrow$ Dropout (0.4) $\rightarrow$ L2 regularization
- Batch normalization
- Output layer (1 unit) with linear activation for regression

Output Feature

The target output, `disk_speed`, was treated as a continuous variable. The model was trained using mean squared error (MSE) as the loss function, with R^2 (coefficient of determination) and mean absolute error (MAE) as evaluation metrics. This configuration was designed to balance model capacity and regularization, ensuring stable convergence while minimizing the risk of overfitting. The inclusion of batch normalization after each dense layer contributed to training stability and improved gradient flow. This architecture allowed the model to process spatial features from both camera perspectives while incorporating the influence of carrier gas flow as contextual information. The use of parallel image encoders enabled the network to

capture complementary visual cues from the powder cone (side view) and the interaction zone (coaxial view), improving its ability to estimate powder feed behavior across a variety of EHLA conditions.

3.4.2 Configuration File (YAML)

The model architecture, data preprocessing, training parameters, and evaluation settings were implemented using the Ludwig AI framework through a structured YAML configuration file. This declarative approach allowed for flexible model design and reproducibility, while abstracting implementation-level details.

The main blocks of the YAML configuration file are described below.

Input Block

Three input features were defined:

- `carrier_gas`: a numerical scalar representing the carrier gas flow in NLPM. This value is passed directly into the model without further transformation.
- `side_path` and `coaxial_path`: file paths to grayscale images captured from side-view and coaxial perspectives, respectively. Although the original images are single-channel (1-channel), they are automatically replicated across three channels by Ludwig to match the input requirements of the EfficientNet-B1 encoder, which is pretrained on ImageNet using RGB images.

Both image inputs were processed by separate EfficientNet-B1 encoders, pretrained on ImageNet and fine-tuned during training.

To improve generalization and robustness, data augmentation was applied during training using Ludwig's built-in augmenters. The following transformations were randomly applied with a 50% probability:

- Horizontal flip
- Rotation (up to ±10 degrees)
- Brightness variation (±20%)
- Contrast variation (±20%)

These augmentations aimed to simulate natural variation in image appearance due to camera noise, lighting shifts, or powder dispersion inconsistencies, thereby improving the model's ability to generalize across experimental noise and subtle visual differences.

Each image input uses the following configuration:

Code 3.1 – Image input configuration

```yaml
preprocessing:
  channels: 3
encoder:
  type: efficientnet
  use_pretrained: true
  trainable: true
  model_cache_dir: null
  model_variant: b1
augmentation:
  - type: random_horizontal_flip
    probability: 0.5
  - type: random_rotate
    degrees: 10
    probability: 0.5
  - type: random_brightness
    max_delta: 0.2
    probability: 0.5
  - type: random_contrast
    max_delta: 0.2
    probability: 0.5
```

EfficientNet-B1 was chosen as the encoder due to its favorable trade-off between accuracy and computational cost. The encoders are pretrained on ImageNet but were fully fine-tuned on the EHLA dataset.

Combiner Block

The outputs of the image encoders and the numerical input were concatenated using a concat combiner. The combined feature vector was processed through two fully connected layers, each with 256 units, ReLU activation, and dropout (0.5) for regularization, as defined in the `num_fc_layers`, `output_size`, and dropout parameters, as shown in the configuration snippet below.

Code 3.2 – Combiner configuration

```yaml
combiner:
 type: concat
 num_fc_layers: 2
 output_size: 256
 dropout: 0.5
 activation: relu
```

Output Block

A single numerical output feature was defined:

Code 3.3 – Output configuration

```yaml
name: disk_speed
type: numerical
preprocessing:
  normalization: zscore
loss:
  type: mean_squared_error
```

```
metrics:
  - type: r2
  - type: mae
```

The output feature `disk_speed` was defined as a numerical target, normalized using z-score preprocessing. The loss function used was mean squared error (MSE), and performance was evaluated using the coefficient of determination (R^2) and mean absolute error (MAE).

The output feature was further processed by a dense regression head composed of multiple fully connected layers with the following structure:

Code 3.4 – Dense regression head configuration

```
layers:
  - type: dense
    num_units: 512
    activation: relu
    dropout: 0.4
    regularizer: l2
    l2: 0.01
  - type: batch_norm
  - type: dense
    num_units: 256
    activation: relu
    dropout: 0.4
    regularizer: l2
    l2: 0.01
  - type: batch_norm
  - type: dense
    num_units: 128
    activation: relu
    dropout: 0.4
    regularizer: l2
    l2: 0.01
  - type: batch_norm
  - type: output
    num_units: 1
    activation: linear
```

The regression head was composed of three fully connected layers with 512, 256, and 128 units, respectively. Each layer included ReLU activation, dropout (0.4), and L2 regularization. Batch normalization was applied after each layer. The final output layer consisted of a single neuron with linear activation for continuous prediction.

Training Configuration Block

The training configuration was optimized for performance and stability under GPU memory constraints (8 GB). Key settings included:

Code 3.5 – Training configuration

```
training:
  batch_size: 32
  epochs: 150
```

```
  early_stop: 15
  learning_rate: 0.0001
  optimizer:
    type: adamw
  learning_rate_scheduler:
    type: reduce_lr_on_plateau
    factor: 0.5
    patience: 5
    min_lr: 0.00001
  use_mixed_precision: true
```

- Batch size and learning rate were selected to balance model convergence and GPU memory usage.
- AdamW was chosen for its regularization benefits.
- A learning rate scheduler was used to reduce the learning rate if validation loss plateaued.
- Early stopping was applied to prevent overfitting.
- Mixed precision training was enabled to accelerate training and reduce memory consumption.

3.4.3 Training Procedure

The training process was conducted using the Ludwig AI framework with the configuration detailed in Section 3.4.2. The model was trained on the Week 2 dataset, consisting of approximately 30,000 images (15,000 synchronized pairs) and their corresponding process parameters. The training was executed on a single GPU (RTX 3070, 8 GB VRAM), which imposed a practical limit on batch size and dataset size.

The dataset was randomly split into training, validation, and test sets using Ludwig's built-in splitting functionality, with the following proportions:

- 70% training
- 10% validation
- 20% test

The model was trained for a maximum of 150 epochs, using a batch size of 32 and the AdamW optimizer with a learning rate of 0.0001. A ReduceLROnPlateau scheduler was applied to decrease the learning rate when validation loss plateaued, with a patience of 5 epochs, a reduction factor of 0.5, and a minimum learning rate of 0.00001. To prevent overfitting, early stopping was triggered if validation performance did

not improve for 15 consecutive epochs. Training was accelerated using mixed precision, reducing memory consumption and increasing throughput. All models were trained from scratch using the EfficientNet-B1 encoders preloaded with ImageNet weights and fine-tuned during training.

3.4.4 Model Evaluation

The trained model was evaluated on the held-out test set using multiple quantitative and visual methods to assess its regression performance and generalization capability. Since the objective was to predict disk speed as a continuous value, the primary evaluation metrics were:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Coefficient of Determination (R^2)

These metrics were automatically computed by Ludwig AI during training and evaluation phases. MAE and RMSE provided direct insight into average and penalized prediction errors, respectively, while R^2 measured how well the predicted values explained the variance in the ground truth.

Beyond scalar metrics, a scatter plot of predicted vs. actual disk speed values was used to visually inspect model performance. Ideally, predictions should align along the diagonal line ($y = x$), indicating high accuracy across the full range of disk speeds. Deviations from this diagonal highlighted underestimation or overestimation tendencies.

The model exhibited stable convergence and no significant signs of overfitting, as validation loss closely tracked training loss throughout the training epochs. Early stopping ensured that training was halted before performance began to degrade. Minor prediction variance was observed at the lower end of the disk speed range, where powder visibility was reduced in the images, potentially impacting feature extraction reliability.

Overall, the evaluation results confirmed that the model was capable of accurately predicting disk speed based on combined visual and process data, even across different carrier gas flow rates.

4 Results

This chapter presents the results obtained from multiple CNN models designed to predict disk speed in the EHLA process based on visual and process inputs. The main model was trained on a balanced dataset from Week 2, and its performance is evaluated using standard regression metrics (MAE, MSE, RMSE, and R^2), as well as visual analyses comparing predicted and actual disk speed values (4.2). In addition to numerical evaluation, Grad-CAM visualizations are used to interpret which regions of the input images (side and coaxial) contributed most to the model's predictions, providing insight into the model's decision-making process (4.5). To support extended interpolation analysis, a second dataset, referred to as Week 3, was collected later in the study. It included intermediate disk speed steps not present in Week 2, such as 5%, 6%, 8%, 9%, 11%, 12%, 14%, and 15% (4.3.2). Although efforts were made to replicate the original acquisition conditions, slight differences in camera alignment and lighting may have introduced distributional shifts between the two datasets. These effects are addressed in the corresponding analysis section (4.3.3). To further assess the model's generalization and robustness, several complementary experiments were conducted. First, a variant of the main model was trained with step 10 removed to evaluate its interpolation capacity when a central value was excluded from training (4.3.1). Second, a combined dataset from Week 2 and Week 3 was used to train an extended model covering all intermediate steps from 4% to 16% in 1% increments (4.3.2). Third, an interpolation test was repeated using this extended dataset, now removing steps 6%, 10%, and 14%, to examine interpolation across a denser input space (4.3.3). Finally, five input configurations of the main model were tested to assess the contribution of each modality (side image, coaxial image, and carrier gas flow) to model accuracy and efficiency (4.4).

4.1 Trained Models and Configuration

The primary model used in this study is referred to as `500_efficientnet_b1_aug_balanced_2`, reflecting its structure and training approach. This architecture is based on EfficientNet-B1 encoders and follows a dual-branch design: one encoder processes side-view images, while the other handles coaxial images. A total of 30,000 images (15,000 synchronized pairs) from Week 2 were used for training, with 3,000 image pairs per disk speed level, covering the range from 4% to 16% in 3% increments. The dataset was carefully balanced to ensure uniform class representation and avoid bias.

To explore interpolation performance, a variation of this model was trained excluding step 10%. This allowed the analysis of how the model predicts an unseen

intermediate value within the central region of the training distribution. To enhance interpolation capability, a new model(`500_efficientnet_b1_aug_balanced_week2_week3`) was trained on a merged dataset combining Week 2 and Week 3. This dataset included 65,000 images (32,500 synchronized pairs) with disk speed values from 4% to 16% in 1% increments. The resulting model was used for full-range evaluation with denser target coverage.

Additionally, a final variant (`week2_week3_no61014`) was trained on the combined dataset, but with steps 6%, 10%, and 14% excluded. This version was used to assess the model's ability to interpolate between densely sampled but deliberately omitted points and to analyze the impact of dataset partitioning on generalization.

Lastly, five alternative input configurations of the main model were evaluated, modifying the input structure to isolate the contribution of each input feature and their combinations. These variants were compared based on prediction accuracy (MAE in z-score units and real units) and inference time, and are summarized in Table 2. This comparison enabled a better understanding of the impact of each data modality on prediction accuracy and computational efficiency, and helped support the choice of the final model.

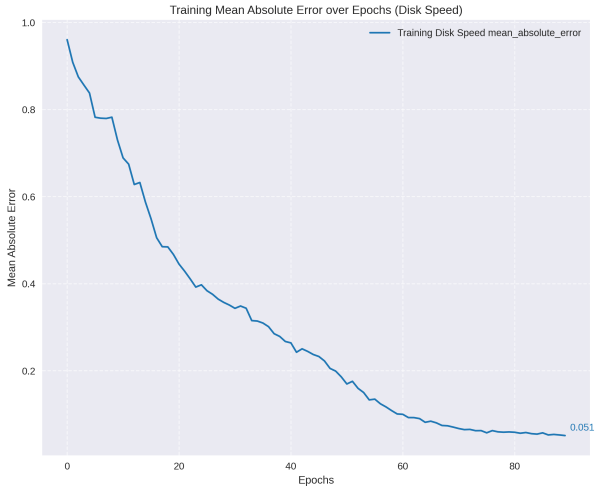
4.2 Quantitative Evaluation

4.2.1 Training and Validation Metrics

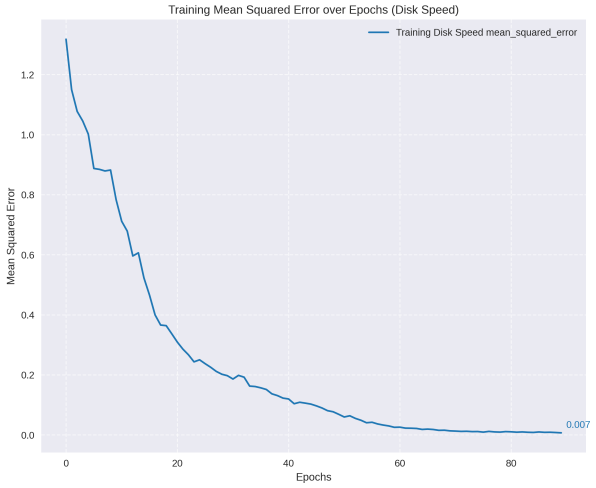
The training of the CNN model was monitored using standard regression metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). Figure 21 presents the evolution of these metrics over 90 training epochs.

As shown in subfigures 21a to 21d, the model exhibited a consistent and smooth learning pattern. The training MAE decreased steadily, reaching a final value of 0.051 in normalized units (approximately 0.22% disk speed error). Similarly, the training MSE and RMSE followed a downward trend, stabilizing at 0.007 and 0.084, respectively. The R^2 score increased progressively and stabilized at 0.993, indicating an excellent fit to the training data.

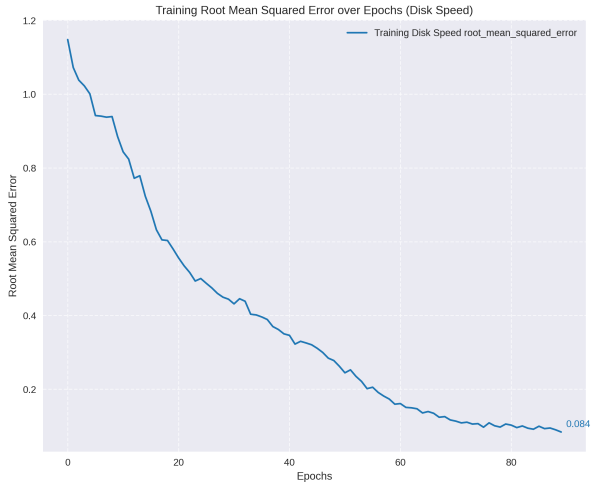
Figure 21 – Training metrics over epochs for disk speed prediction: (a) MAE, (b) MSE, (c) RMSE, and (d) R^2.



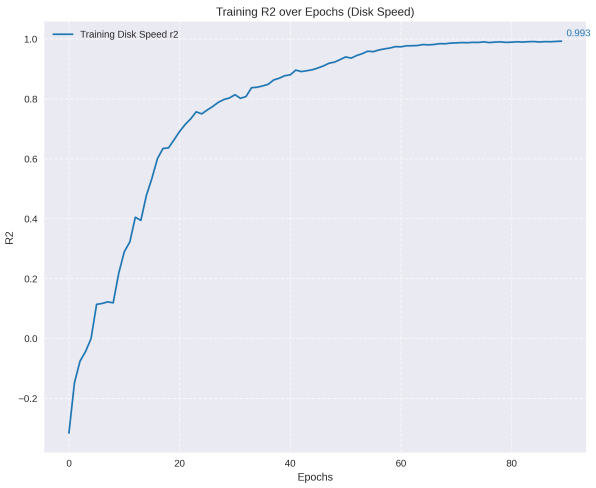
(a) Training MAE over Epochs



(b) Training MSE over Epochs



(c) Training RMSE over Epochs

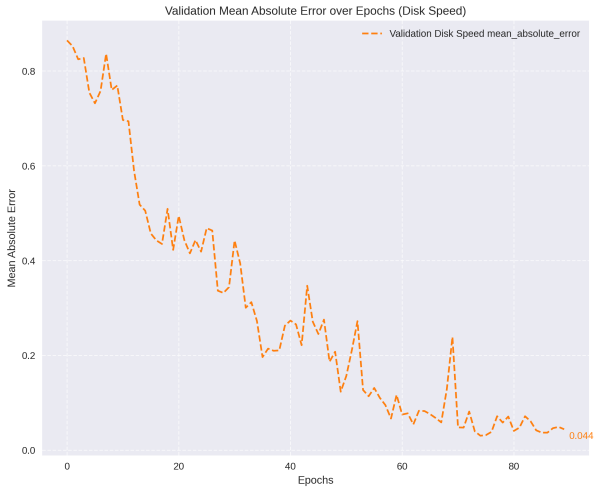


(d) Training R^2 over Epochs

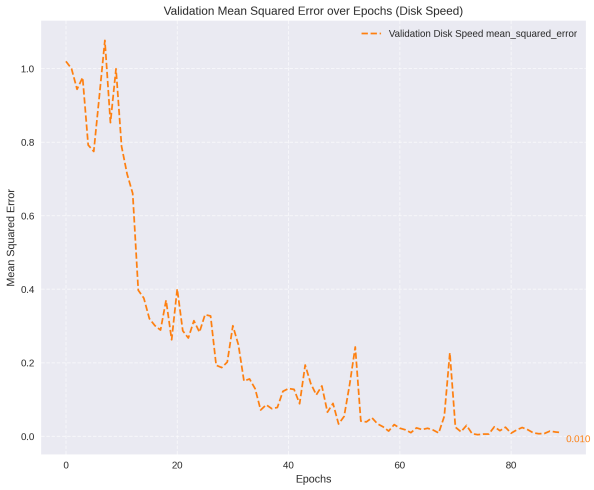
Source: The Author.

Validation performance, presented in Figure 22, shows a similar positive trend to the training phase, although with slightly more variation due to data complexity. As shown in subfigures 22a to 22d, the validation MAE reached 0.044 (approximately 0.19% disk speed error) by the final epoch, while the MSE and RMSE dropped to 0.010 and 0.102, respectively. The validation R^2 score fluctuated around 0.980, confirming the model's strong generalization capability throughout the training process.

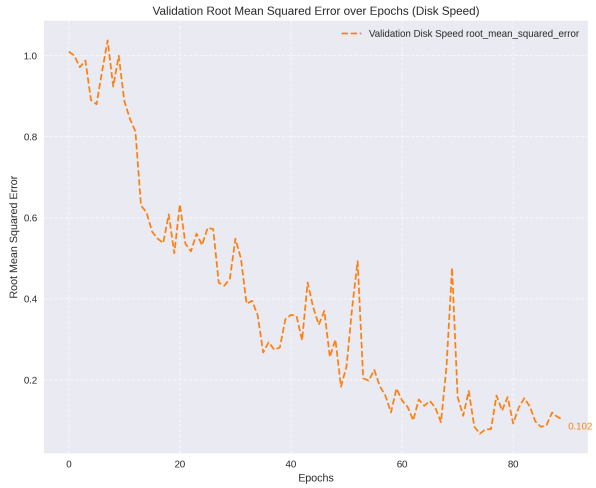
Figure 22 – Validation metrics over training epochs for disk speed prediction: (a) MAE, (b) MSE, (c) RMSE, and (d) R².



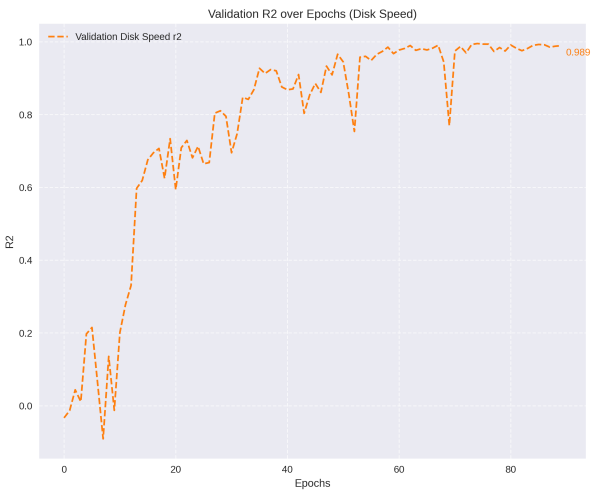
(a) Validation MAE over Epochs



(b) Validation MSE over Epochs



(c) Validation RMSE over Epochs



(d) Validation R² over Epochs

Source: The Author.

A direct comparison of training and validation loss curves is depicted in Figure 23. The curves remain closely aligned throughout training, with no signs of overfitting. Small oscillations in the validation loss reflect the presence of local variations in the data but do not indicate instability.

Figure 23 – Training vs Validation Loss

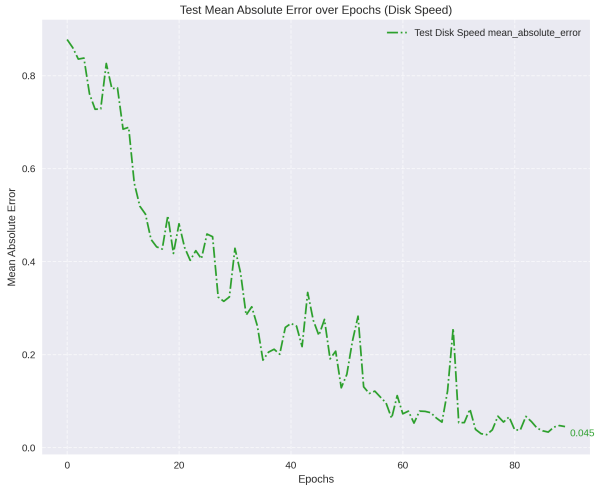
Source: The Author.

Overall, the results demonstrate stable convergence and high predictive power, with consistent performance across training and validation datasets.

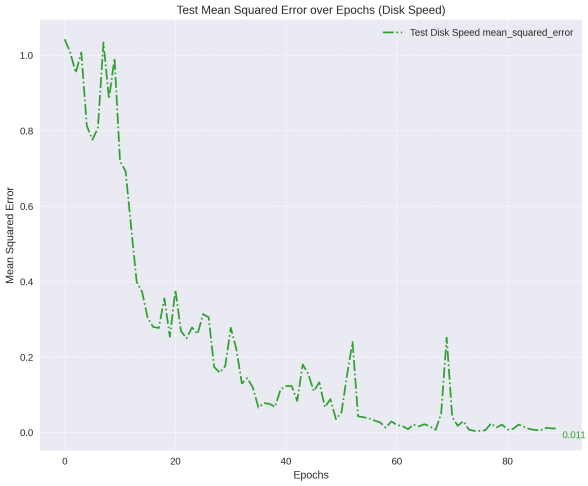
4.2.2 Performance on the Test Set

The model's performance on the held-out test set was evaluated using the same metrics applied during training and validation. The test results confirm the model's ability to generalize to unseen data, with consistently low errors and high correlation.

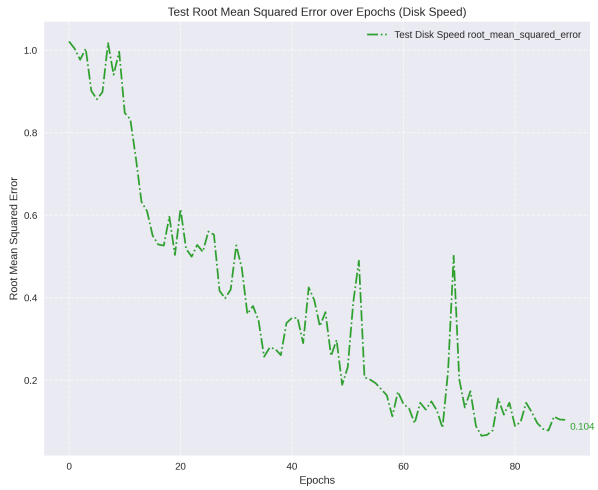
Figure 24 – Evaluation metrics on the test set across training epochs: (a) MAE, (b) MSE, (c) RMSE, and (d) R^2.



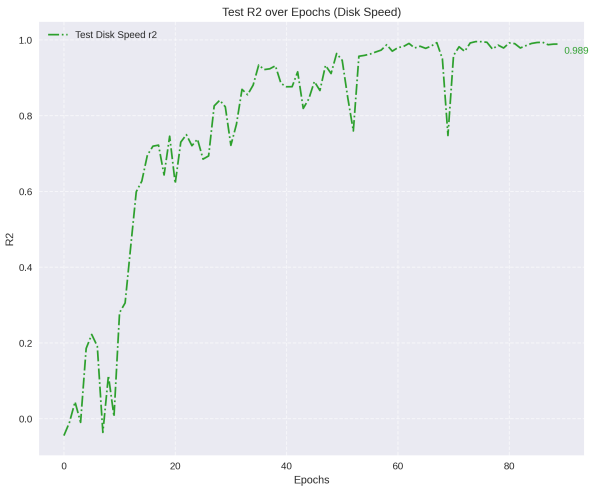
(a) Test MAE over Epochs



(b) Test MSE over Epochs



(c) Test RMSE over Epochs



(d) Test R^2 over Epochs

Source: The Author.

As shown in Figures 24a to 24d, the model maintained strong performance on the unseen test set. The MAE reached a final value of 0.045 in z-score units, which corresponds to approximately 0.19% disk speed error, considering a standard deviation of 4.2427. The MSE and RMSE settled at 0.011 and 0.104, respectively, consistent with the validation results. The R^2 stabilized at 0.989, confirming the model's ability to explain the variance in the test data. The evolution curves across epochs remain smooth, with no signs of overfitting, and fluctuations are limited to natural variations in the test set.

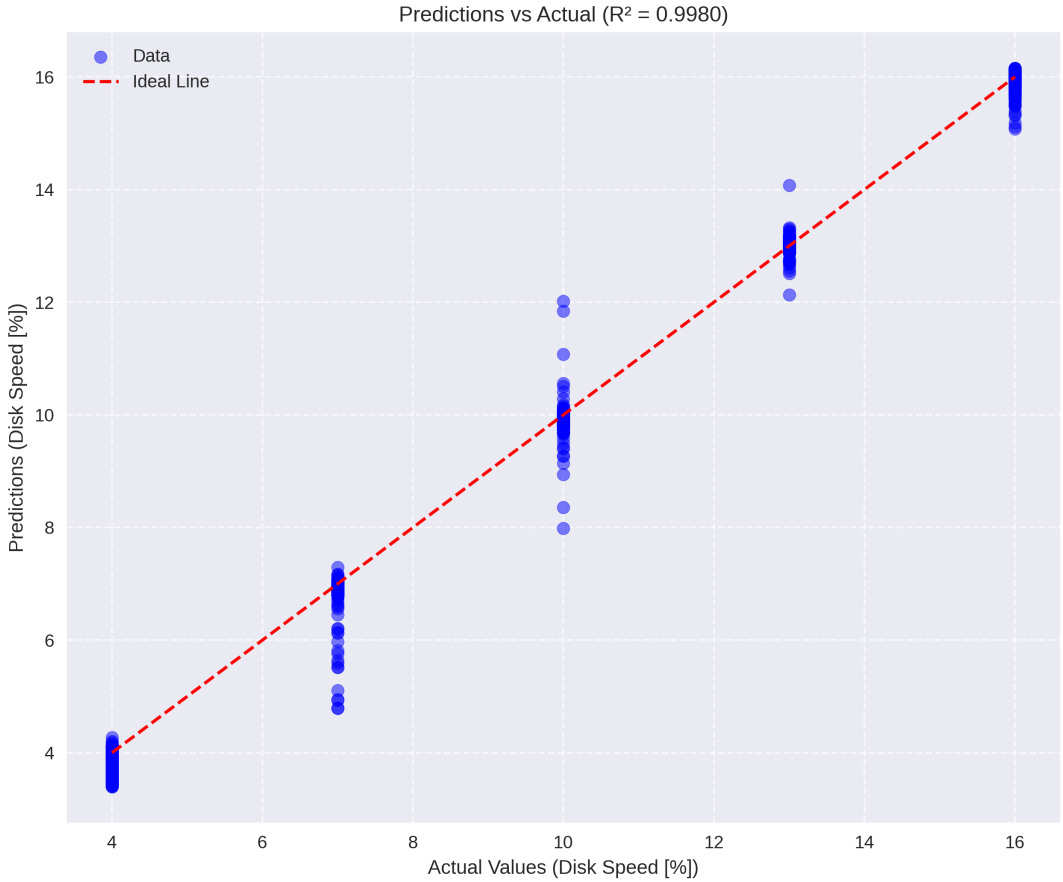
Overall, these metrics indicate strong generalization and robustness of the trained model, validating its applicability for real-time disk speed prediction tasks.

4.3 Visual Evaluation of Predictions

To further assess the model's predictive capabilities, a visual comparison between predicted and actual disk speed values was conducted using several diagnostic plots.

Figure 25 displays the scatter plot of predicted versus actual values for the test set, along with the ideal prediction line (in red). The data points are closely distributed around the diagonal, and the coefficient of determination (R^2) is 0.9980, indicating strong agreement between predicted and true values.

Figure 25 – Predictions vs Actual

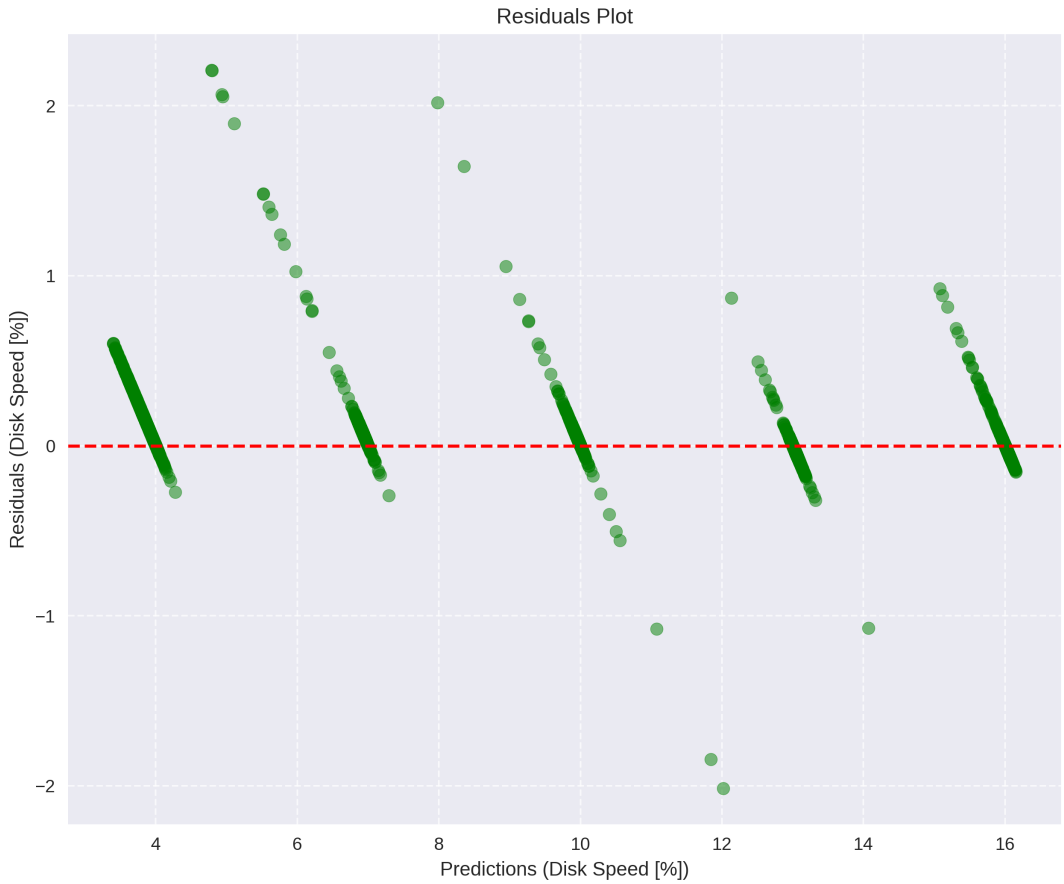


Source: The Author.

The residuals plot in Figure 26 provides a more detailed look at the prediction errors across the prediction range. The majority of the residuals are tightly clustered around zero, forming vertical bands at each discrete disk speed level. Most prediction errors lie within ±0.5 disk speed %, with only a few outliers exceeding ±1 disk speed %. This indicates that the model maintains consistent accuracy across the prediction range, with minor deviations at some intermediate steps. Additionally, some systematic deviations are visible, especially at lower and higher prediction ranges. This could

indicate slight underfitting at intermediate speeds or model bias in transition zones between classes.

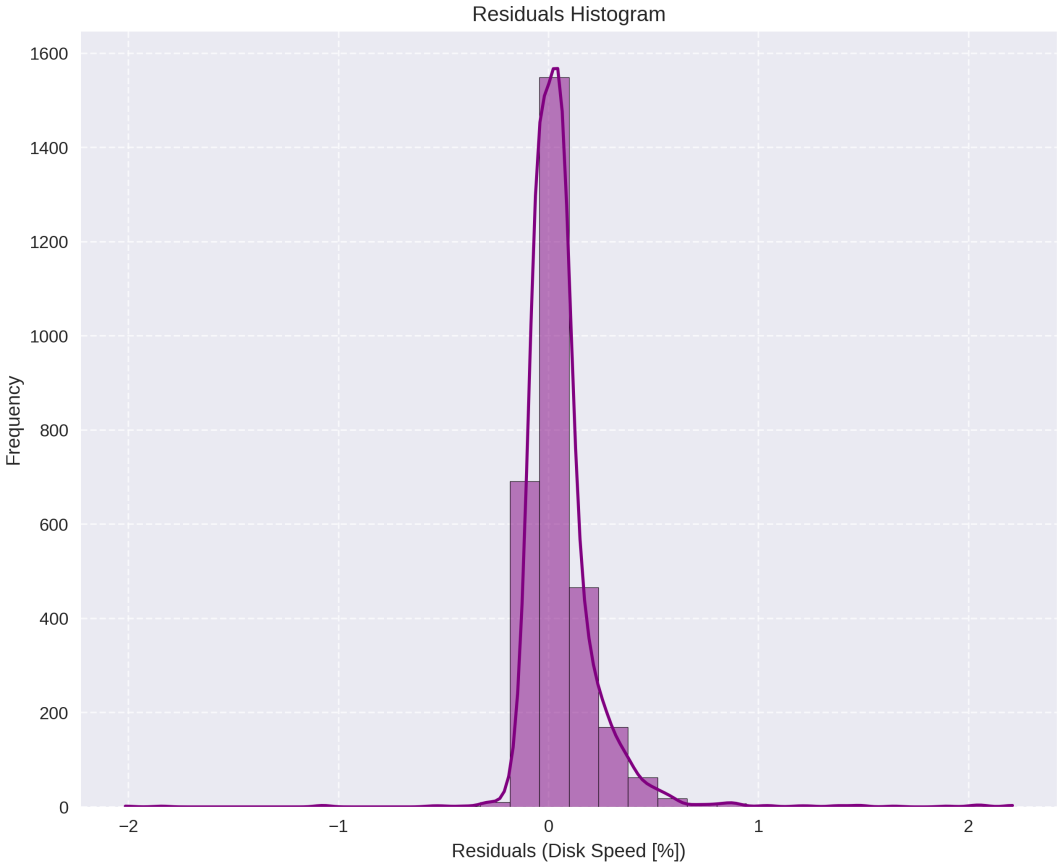
Figure 26 – Residuals Plot



Source: The Author.

The distribution of residuals is shown in Figure 27. The histogram demonstrates a highly concentrated distribution centered around zero, with the majority of errors confined within a narrow band of ±0.5 disk speed %. This reflects high model precision and low variability. The smooth kernel density estimate confirms a unimodal distribution, indicating consistent error behavior.

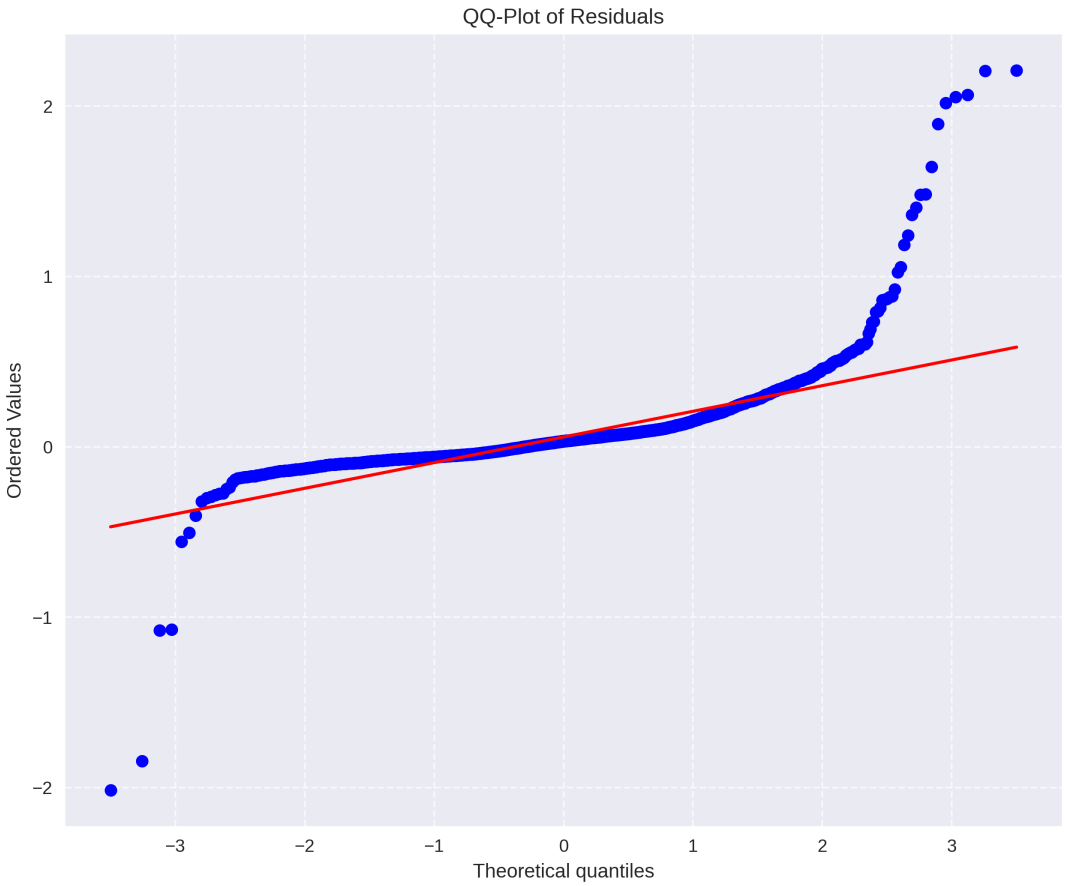
Figure 27 – Residuals Histogram



Source: The Author.

To validate the normality of residuals, Figure 28 presents a Q-Q plot. Although most points align well with the theoretical quantile line, substantial deviations are observed at both tails, especially in the upper quantiles. This suggests that while the residuals are approximately normal in the center, the distribution has heavier tails than a Gaussian. These deviations likely stem from rare cases where the model slightly misestimates at the extreme ends of the disk speed spectrum.

Figure 28 – Quantile-Quantile plot of Residuals



Source: The Author.

Together, these visualizations confirm the model's strong predictive ability and statistical consistency. The residual analysis supports the robustness of the predictions, with only minor deviations from ideal behavior concentrated in specific regions.

4.3.1 Interpolation Evaluation: Step 10 Exclusion

To assess the model's ability to interpolate between unseen values, an experiment was conducted in which all training samples corresponding to disk speed step 10% were excluded. The model was then trained on the remaining values (4%, 7%, 13%, and 16%) and evaluated on a complete test set containing all disk speeds, including the unseen 10% step.

Although the model was trained using a regression objective, the results revealed a behavior typically associated with classification: the predictions tended to cluster around the known training targets, forming discrete bands rather than a smooth distribution.

4.3.1.1 Predicted vs. Actual Plot

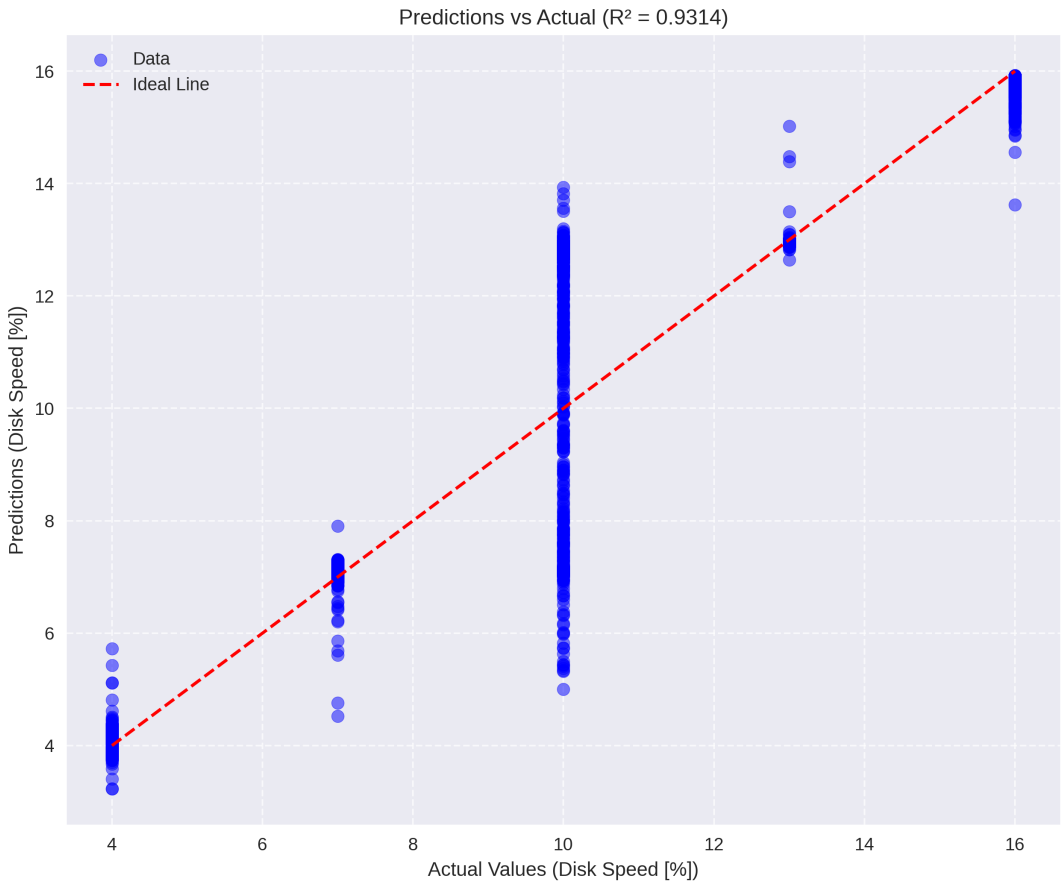
As shown in Figure 29, the predicted values exhibit a clear tendency to group around the trained disk speeds (4%, 7%, 13%, and 16%), even though the model was not trained to classify. For the unseen step 10%, out of 600 test samples (representing one fifth of the dataset), as shown in the residual histogram in Figure 48:

Approximately 450 predictions (~75%) were drawn toward the nearest trained values, predominantly step 7 and step 13, with residuals around ±3 disk speed %. This suggests that the model tends to gravitate toward familiar training points when exposed to novel inputs.

Around 150 predictions (~25%) fell closer to the true value of 10%, indicating some capacity for interpolation based on feature continuity. This behavior reflects the model's partial generalization ability in scenarios where input features exhibit gradual transitions.

This clustering effect highlights a common phenomenon in regression tasks with discrete and sparsely distributed target values: when intermediate steps are missing, the model may generalize poorly and "snap" to the nearest learned outputs.

Figure 29 – Predictions vs Actual for Step 10 (Interpolation Test)

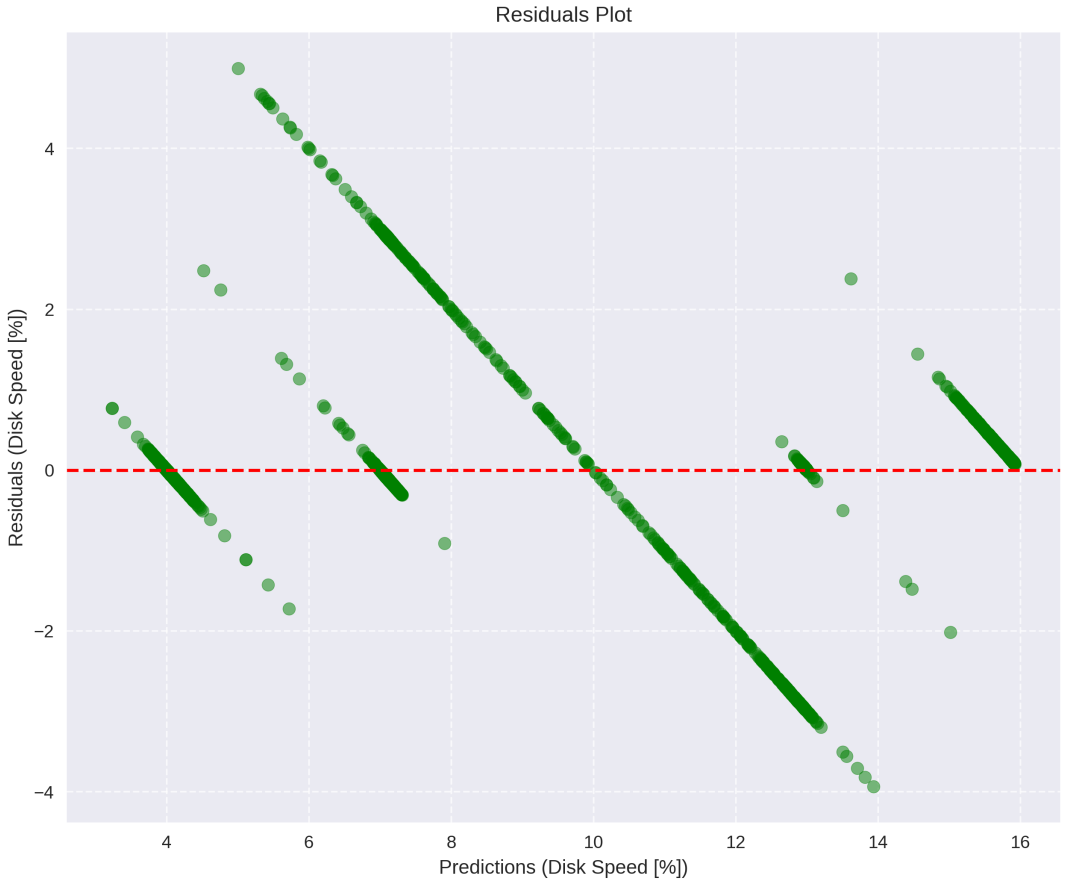


Source: The Author.

4.3.1.2 Residuals Plot

Figure 30 shows the residuals of the predictions. The predictions for the 10% step demonstrate systematic errors, with many samples underestimated or overestimated by ±3 disk speed %, aligning with steps 7 and 13. The residuals span from approximately -4.5 to +3.3 disk speed %, forming two strong directional clusters.

Figure 30 – Residuals Plot for Interpolation Test (All Steps)

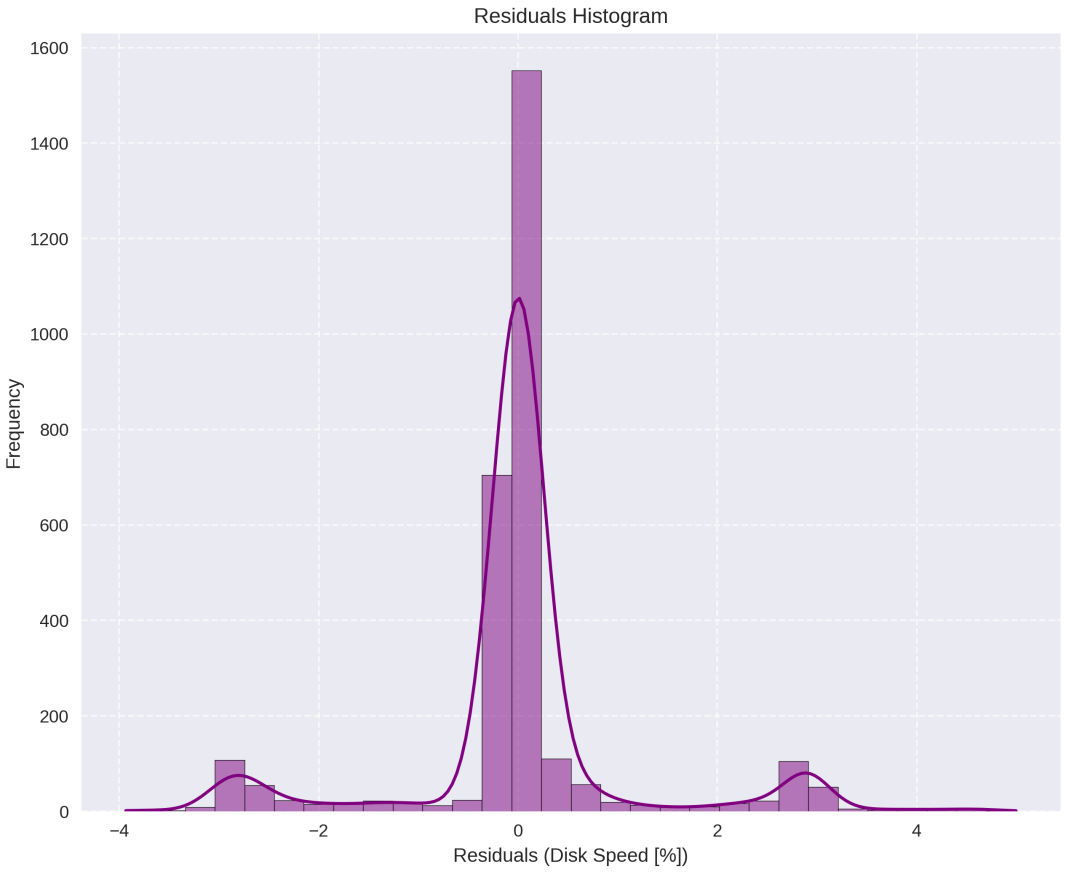


Source: The Author.

4.3.1.3 Residuals Histogram

As shown in Figure 31, the histogram of residuals reveals a tri-modal distribution, with two broad peaks corresponding to the misclassified predictions (toward 7% and 13%) and one central peak corresponding to low-residual predictions, which includes both correctly predicted samples from trained steps as well as the subset of step 10 predictions that were successfully interpolated. This confirms that, although the model regressed values on a continuous scale, it behaved more like a discrete predictor in this case.

Figure 31 – Residuals Histogram for Step 10 Interpolation

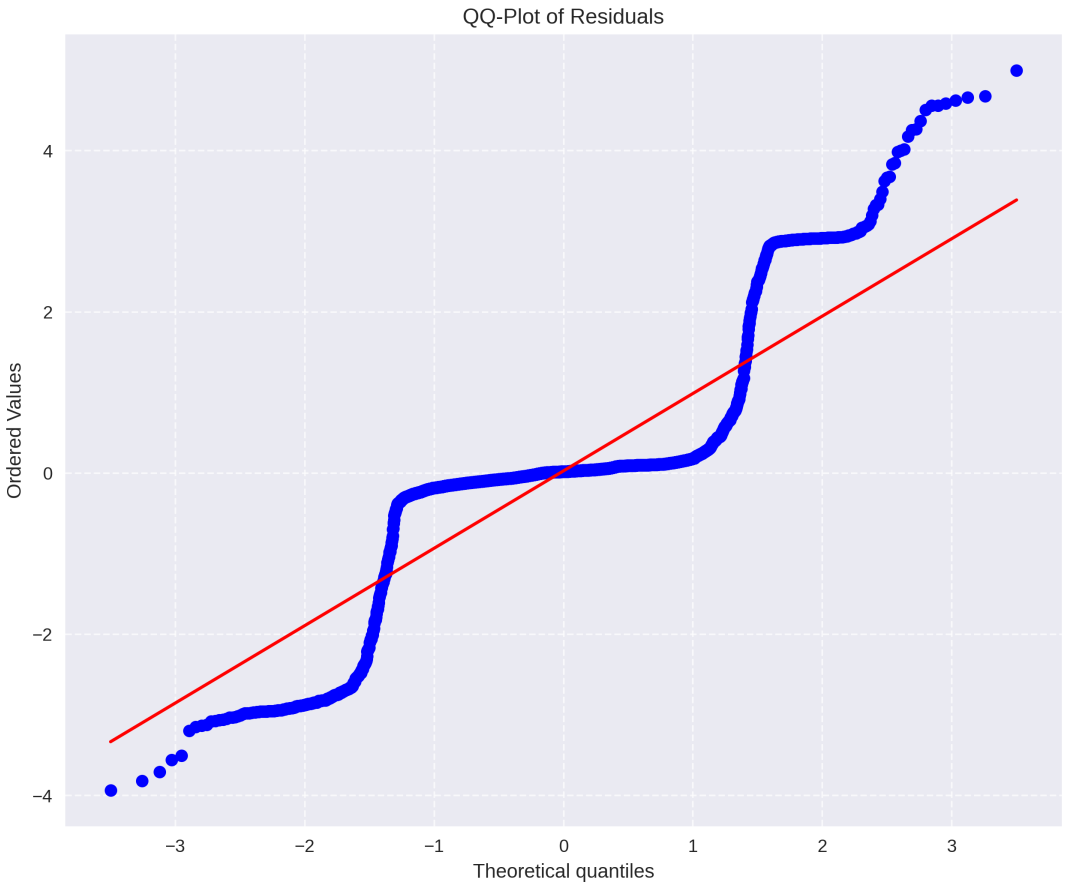


Source: The Author.

4.3.1.4 Q-Q Plot of Residuals

The Q-Q plot in Figure 32 confirms that the residuals do not follow a normal distribution. The strong stepwise shape of the plot indicates discrete grouping in the predictions, which is a direct consequence of the model's limited exposure to intermediate values and its reliance on patterns from the nearest known classes.

Figure 32 – Quantile-Quantile plot of Residuals for Step 10 Interpolation



Source: The Author.

4.3.2 Extra Intermediate points (Week 2 + Week 3)

In order to improve the model's interpolation capability between discrete disk speed levels, a new dataset was assembled by combining the original Week 2 images, which contained only the steps 4, 7, 10, 13, and 16, with additional data acquired in Week 3, targeting the intermediate steps 5, 6, 8, 9, 11, 12, 14, and 15. The final dataset consisted of 65,000 images, forming 32,500 synchronized coaxial/side-view pairs, with disk speed values ranging from 4% to 16% in steps of 1%.

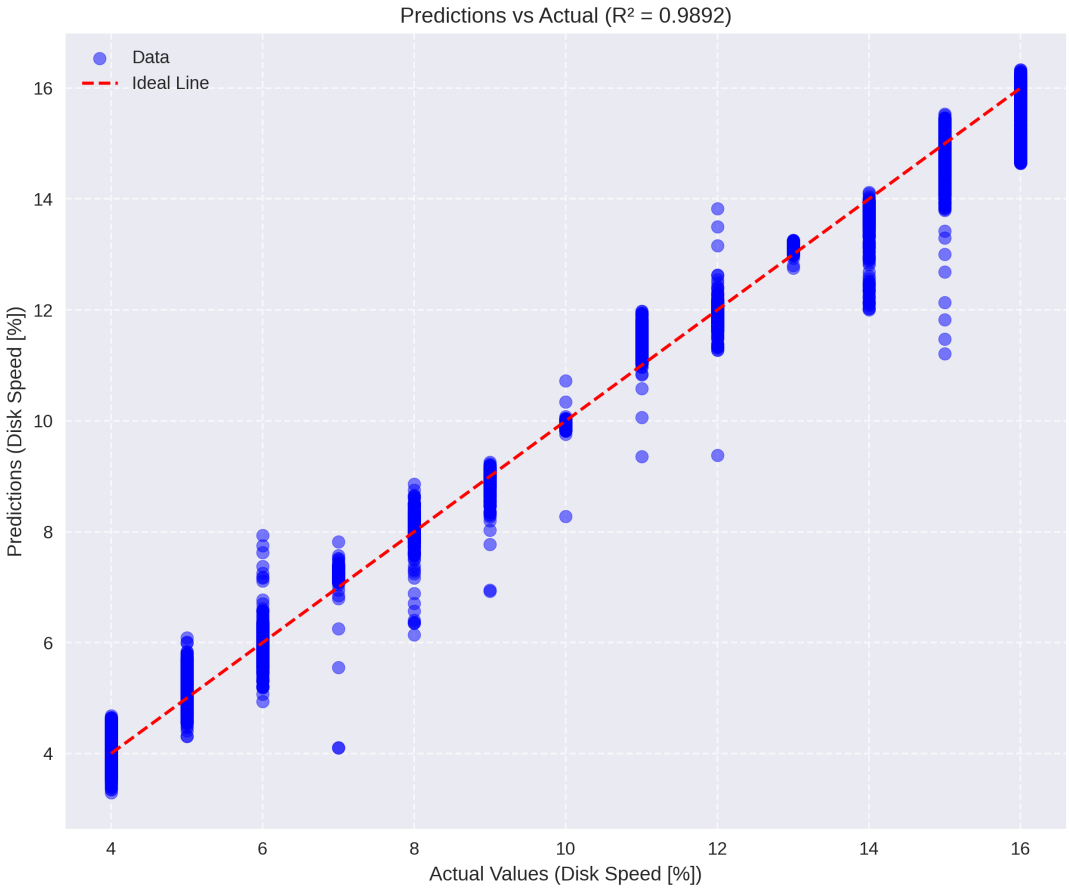
This dataset was used to train a new model, referred to as `500_efficientnet_b1_aug_balanced_week2_week3`, with the goal of encouraging smoother regression behavior and eliminating prediction snapping between discrete levels. However, an important consideration is that the Week 3 data were acquired on a different day, requiring the experimental setup (including camera alignment and lighting) to be manually rebuilt. Although care was taken to match the original conditions, subtle variations in camera aperture and illumination positioning may have introduced distributional shifts between the two datasets.

Despite these inconsistencies, the model achieved good predictive performance on test samples that belonged to the same training distribution. However, further

analysis revealed that the model tended to treat Week 2 and Week 3 regions as separate clusters, rather than generalizing across them as a unified continuum.

Figure 33 shows the scatter plot of predicted vs. actual values, revealing a strong alignment with the ideal line ($R^2 = 0.9892$), indicating high predictive accuracy across all speeds.

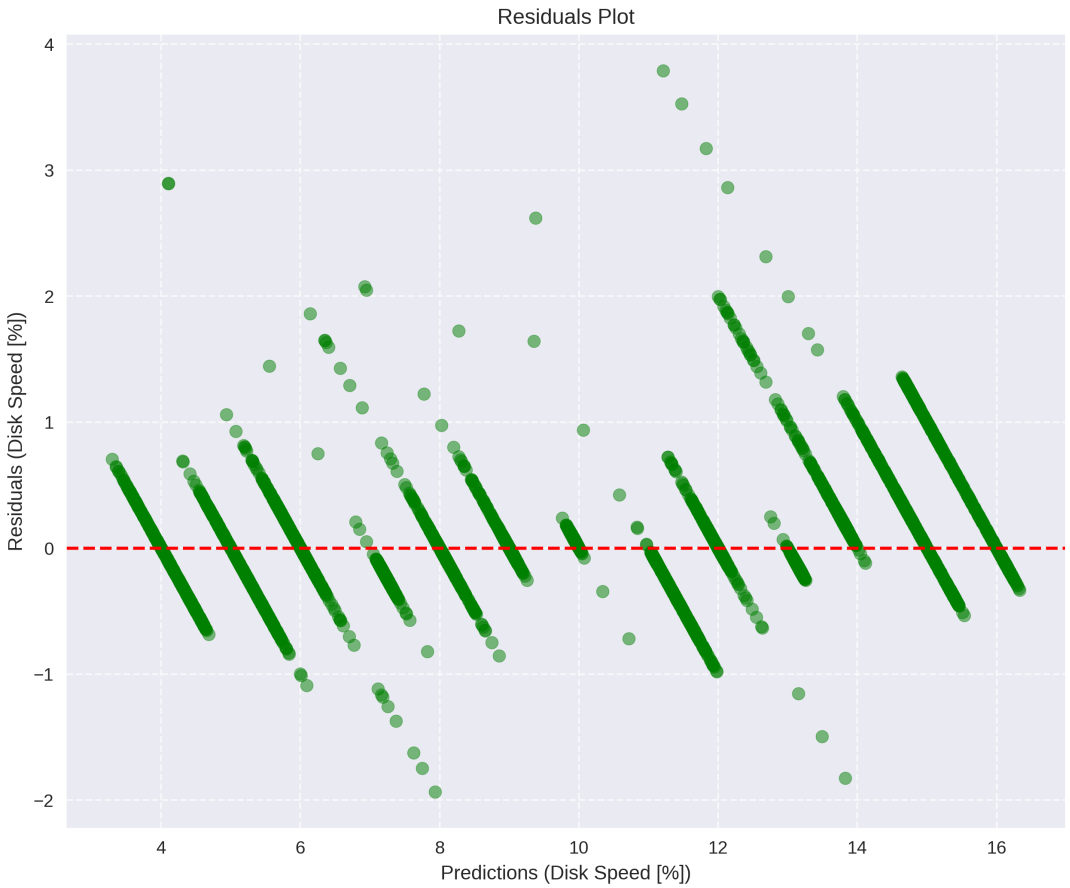
Figure 33 – Predictions vs Actual



Source: The Author.

The corresponding residuals plot in Figure 34 shows tight clusters centered around zero, though with minor asymmetry and a few outliers concentrated in mid-range predictions.

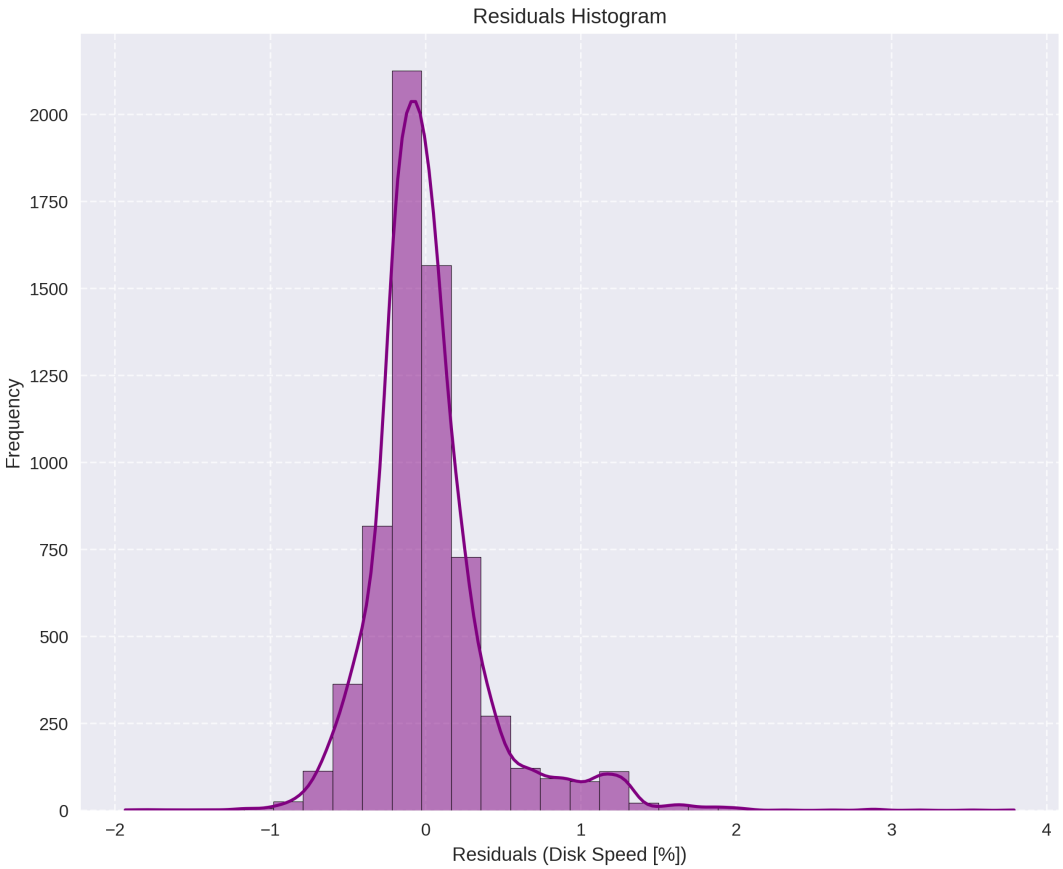
Figure 34 – Residuals Plot



Source: The Author.

Figure 35 presents the residuals histogram, which confirms a narrow error distribution, with most residuals within ±0.5 disk speed %.

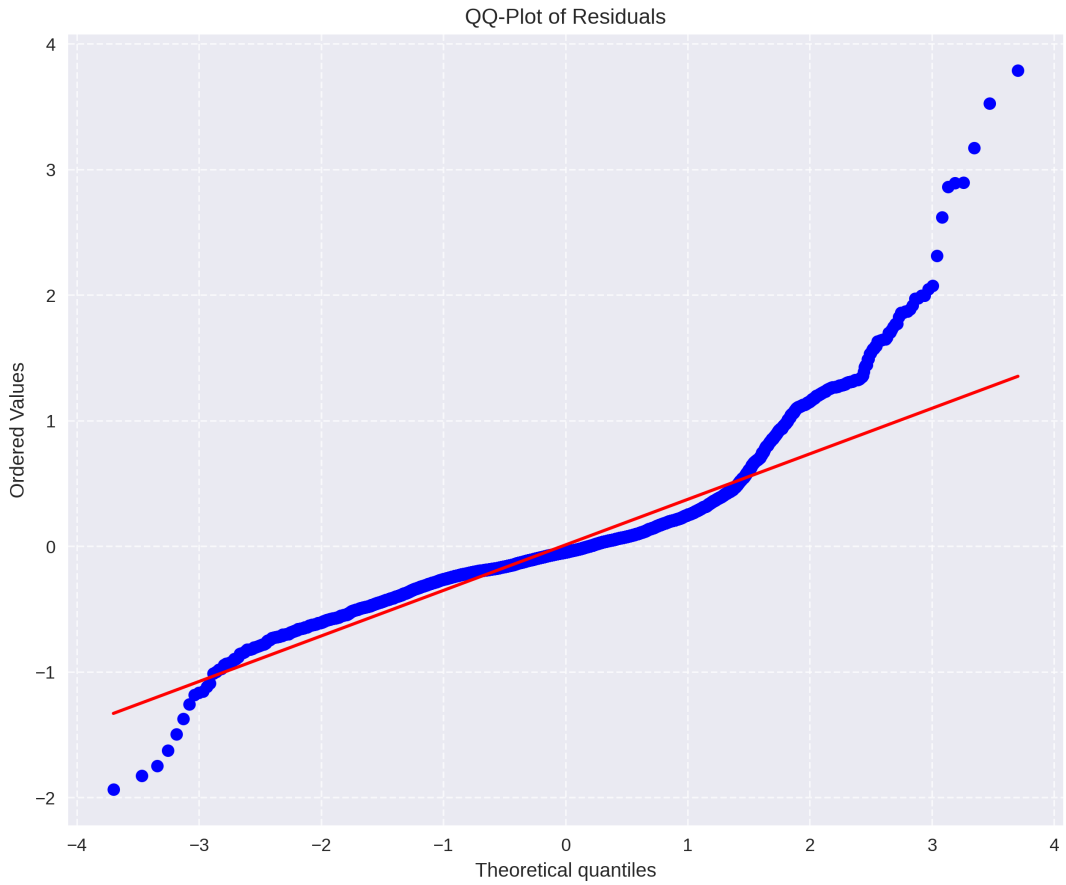
Figure 35 – Residuals Histogram



Source: The Author.

Finally, the QQ-plot in Figure 36 demonstrates that residuals approximate a normal distribution with minor deviations in the tails, especially for extreme predictions.

Figure 36 – Quantile-Quantile plot



Source: The Author.

4.3.3 Interpolation Behavior – Excluding Points 6, 10, and 14

To evaluate the model's interpolation capacity across the extended range, this version, named `500_efficientnet_b1_aug_balanced_week2_week3_no61014_1`, was trained on the same combined dataset as before, but with the deliberate exclusion of disk speed values 6%, 10%, and 14%. The objective was to verify whether the model could generalize to unseen intermediate speeds by leveraging the contextual information from neighboring steps.

Evaluation on the complete test set showed that while the model retained overall accuracy, its interpolation behavior revealed a bias toward the dataset boundaries rather than toward adjacent values. For example:

- When step 6 was excluded, predictions for that point were drawn toward steps 5 and 8, but predominantly toward the step that belonged to the same dataset (e.g., Week 3).
- When step 10 was removed, predictions were attracted to steps 7 or 13 (Week 2), rather than steps 9 or 11 (Week 3).

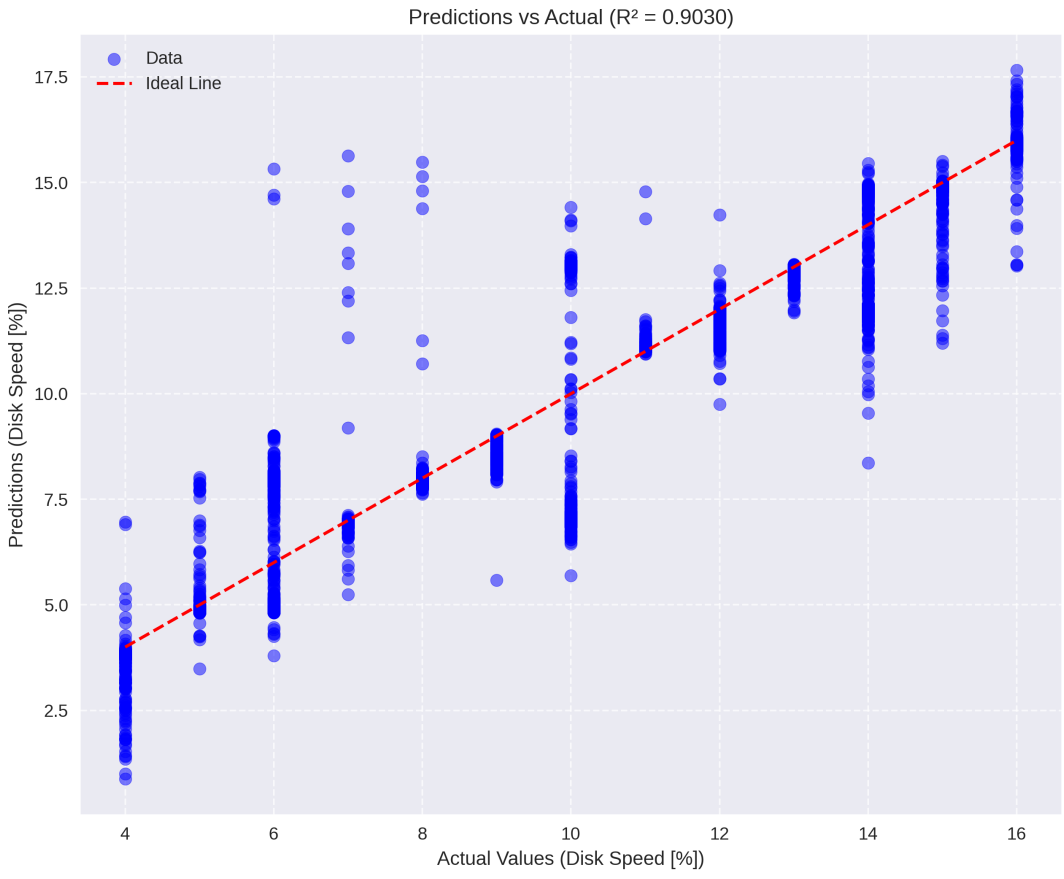
- Similarly, for step 14, predictions were pulled toward 12 or 15, depending on their dataset origin.

These results indicate that although the model was exposed to a full range of disk speeds, the underlying domain shift between Week 2 and Week 3 limited its ability to generalize globally across datasets. Instead, the model learned local continuity within each dataset, resulting in interpolation that favored intra-dataset neighbors over global regression across the full spectrum.

This behavior underscores the importance of maintaining a consistent acquisition setup when aiming for smooth and reliable regression across continuous process parameters.

As shown in Figure 37, the Predicted vs. Actual plot illustrates that the predictions for the excluded steps (6%, 10%, 14%) do not follow a continuous regression trend. Instead, they cluster around neighboring values that belong to the same dataset partition (Week 2 or Week 3), revealing dataset-dependent interpolation patterns. This suggests that the model has learned separate local mappings for each acquisition group rather than forming a truly global regression function.

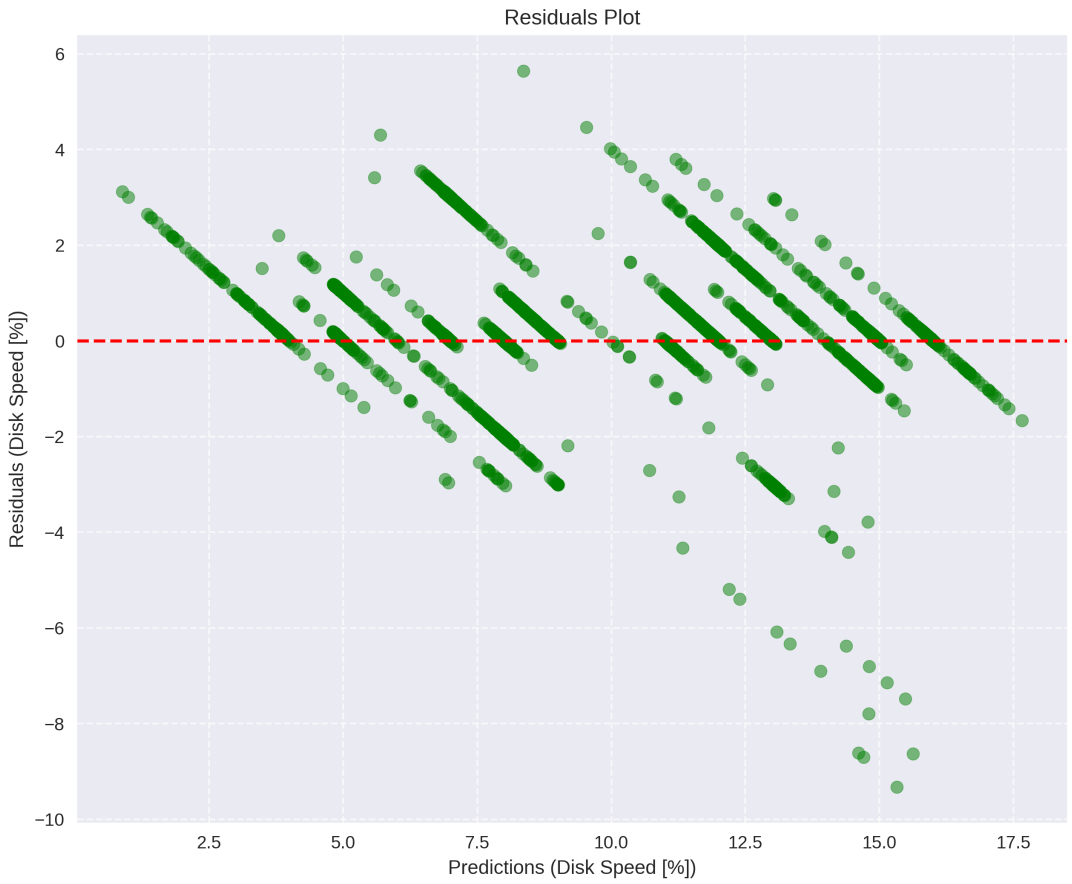
Figure 37 – Predicted vs. Actual Plot



Source: The Author.

This behavior becomes even more evident in the Residuals Plot (Figure 38), where the residuals for excluded points form distinct vertical structures. The errors are systematically biased toward the training points in the same acquisition group, confirming the lack of a smooth transition between datasets in the learned representation.

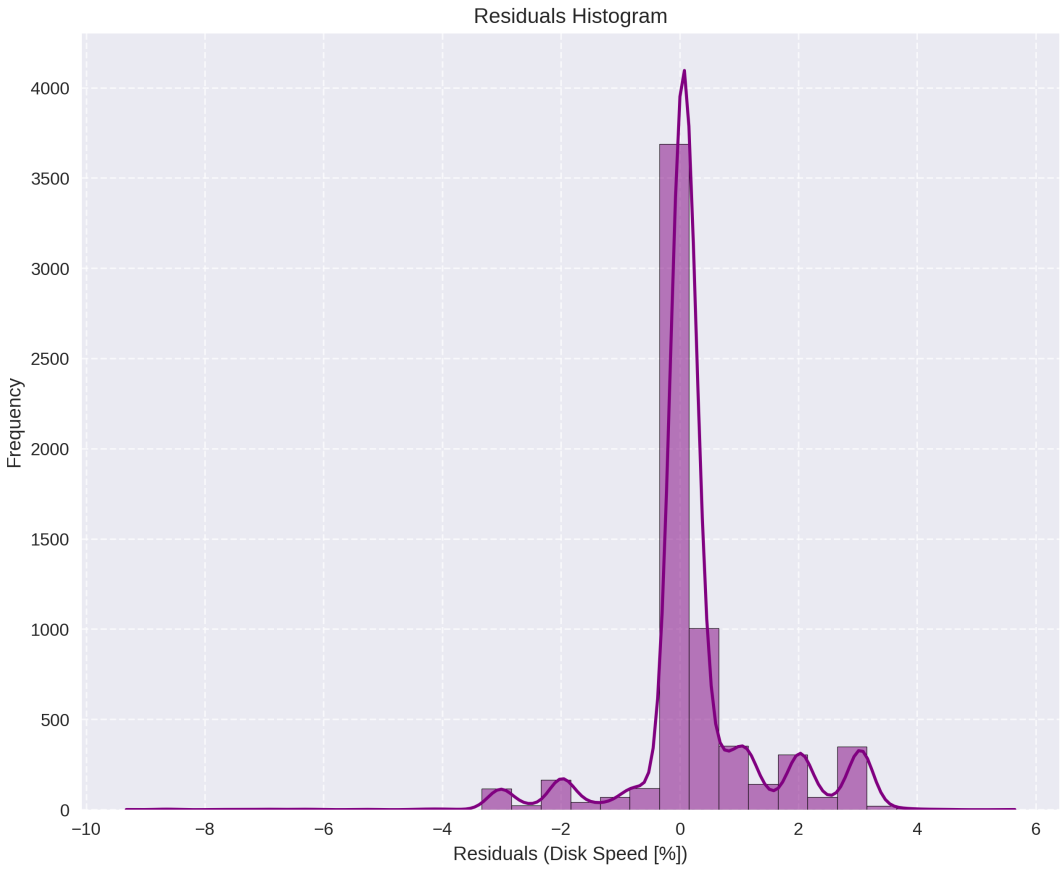
Figure 38 – Residuals Plot



Source: The Author.

Further support for this pattern is seen in the Residuals Histogram (Figure 39), where multiple residual modes appear. Significant peaks correspond to under- and overestimations of the excluded points, reinforcing the presence of prediction snapping and confirming that residuals are not symmetrically centered around zero.

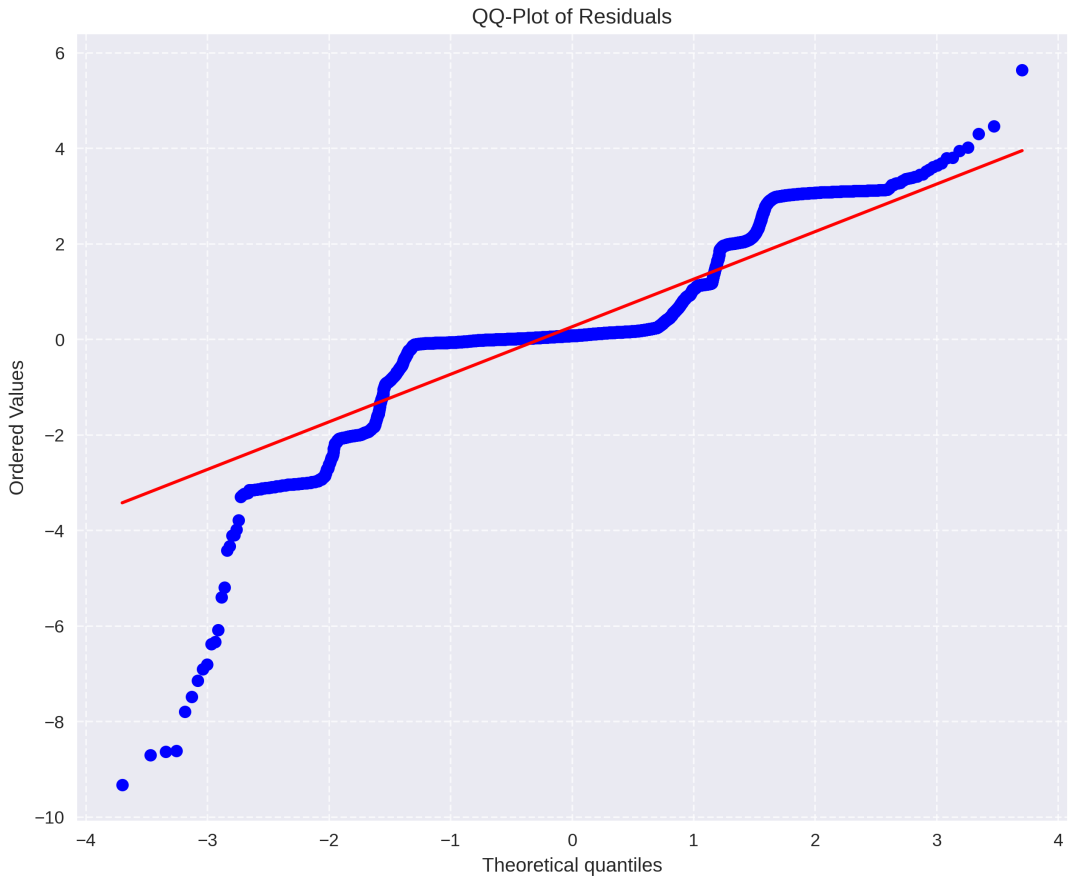
Figure 39 – Residuals Histogram



Source: The Author.

Finally, the Q-Q Plot (Figure 40) presents a strong deviation from normality, with visible steps and asymmetries in the tails. These distortions indicate that the residuals for the excluded steps are not smoothly distributed, reflecting the model's limited global generalization despite full numeric coverage in the training range.

Figure 40 – Q-Q Plot



Source: The Author.

4.4 Model Comparison

To assess the influence of different input combinations on model performance, five variants of the best-performing architecture were trained using different subsets of inputs: side image, coaxial image, and carrier gas flow. Table 2 summarizes the results, showing the number of features, mean absolute error (MAE) in z-score units, and the average time required for a single prediction. Model 1, which combines all three inputs (side image, coaxial image, and carrier gas), achieved a balanced performance with a low MAE of 0.029, equivalent to approximately 0.12% disk speed error, while maintaining a prediction time of 9.271 ms. Model 3, which uses only the side image, achieved the lowest MAE of 0.014 (~0.06% disk speed error) and a significantly faster inference time (5.310 ms), suggesting that side images alone carry strong predictive information.

In contrast, Model 5, which uses only the carrier gas input, performed poorly with an MAE of 0.857, corresponding to a disk speed error of approximately 3.64%, with the lowest computational cost (0.630 ms). This underscores its limited predictive value when used in isolation.

These results indicate that visual information, particularly from side images, plays a crucial role in model accuracy. The inclusion of coaxial images and carrier gas data can enhance interpretability and robustness but may increase inference time.

Therefore, there is a clear trade-off between prediction accuracy and computational efficiency, depending on the input configuration.

Table 2 – Model Evaluation Based on Input Combinations: MAE and Prediction Time Across Five Model Variants

Model	Side Image	Coaxial Image	Carrier Gas	Number of Features	MAE zscore	MAE Disk speed (%)	Time for one prediction (ms)
1	X	X	X	3	0.029	0.123	9.271
2	X	X		2	0.107	0.453	9.124
3	X			1	0.014	0.059	5.310
4		X		1	0.154	0.652	5.402
5			X	1	0.857	3.633	0.630

Source: The Author.

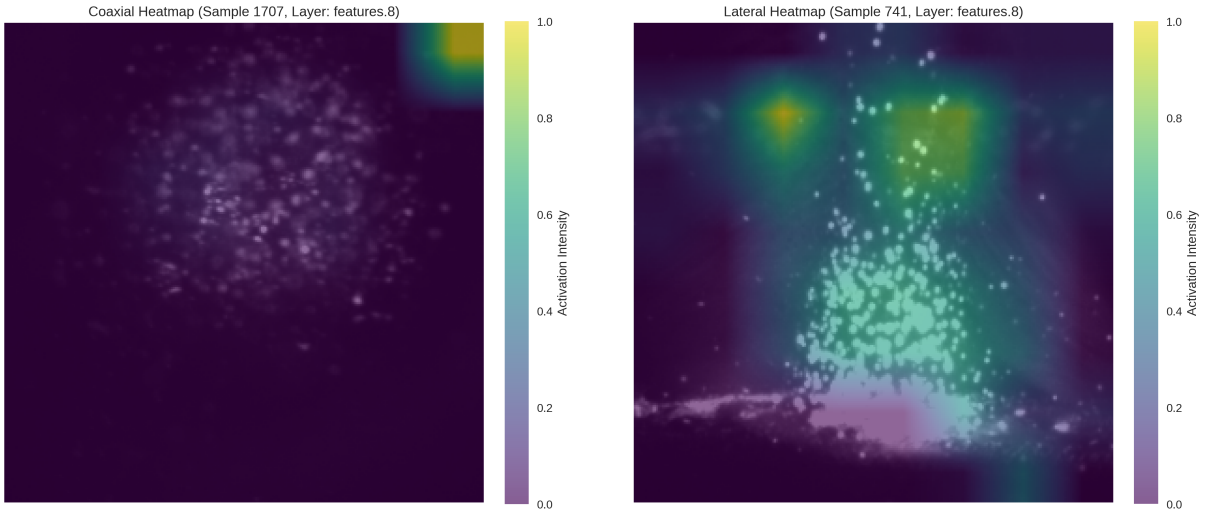
4.5 Interpretability Analysis with Grad-CAM

To better understand how the model makes predictions from visual data, the Grad-CAM (Gradient-weighted Class Activation Mapping) technique was applied to selected convolutional layers in both image branches of the model. Grad-CAM uses the gradients of the target output (in this case, the predicted disk speed) flowing into the final convolutional layers to produce heatmaps, highlighting the most relevant regions of the input image that influenced the prediction.

In this study, two independent EfficientNet-B1 encoders were used: one for the side-view images and one for the coaxial images. Grad-CAM was computed for feature layers ranging from features.0 to features.8, with visualizations presented in Figures 41–43 for the higher layers (features.6 to features.8), where activations become more semantically meaningful.

For the side images, the activation maps show strong and consistent focus on key process regions such as the powder stream, melt pool, and the plume above the interaction zone. These regions are visible as high-activation zones centered vertically in the image and correlate well with physical phenomena associated with changes in disk speed.

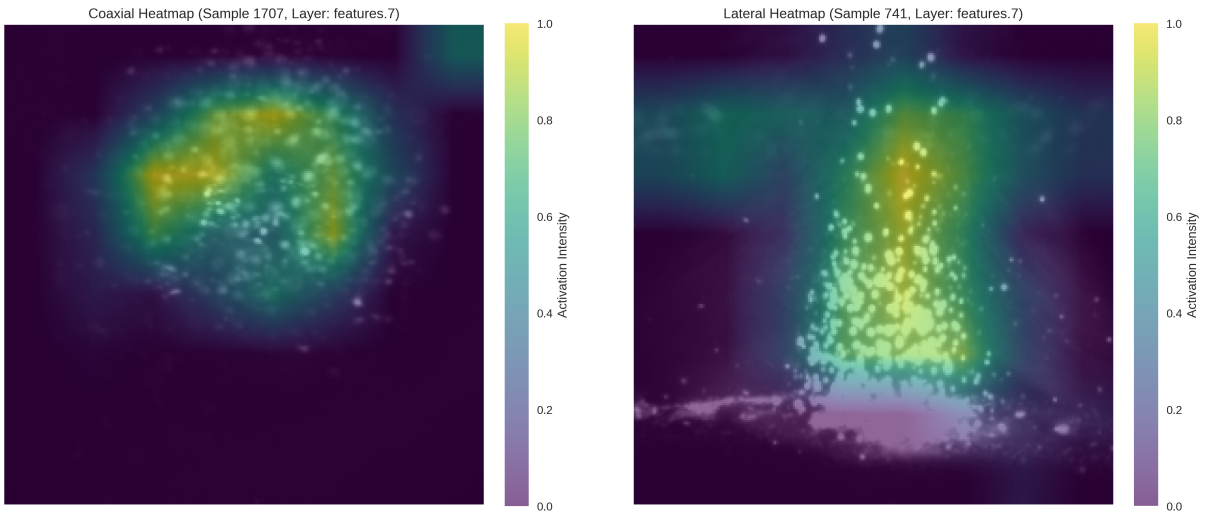
Figure 41 – Grad-Cam Heatmap - Feature 8



Source: The Author.

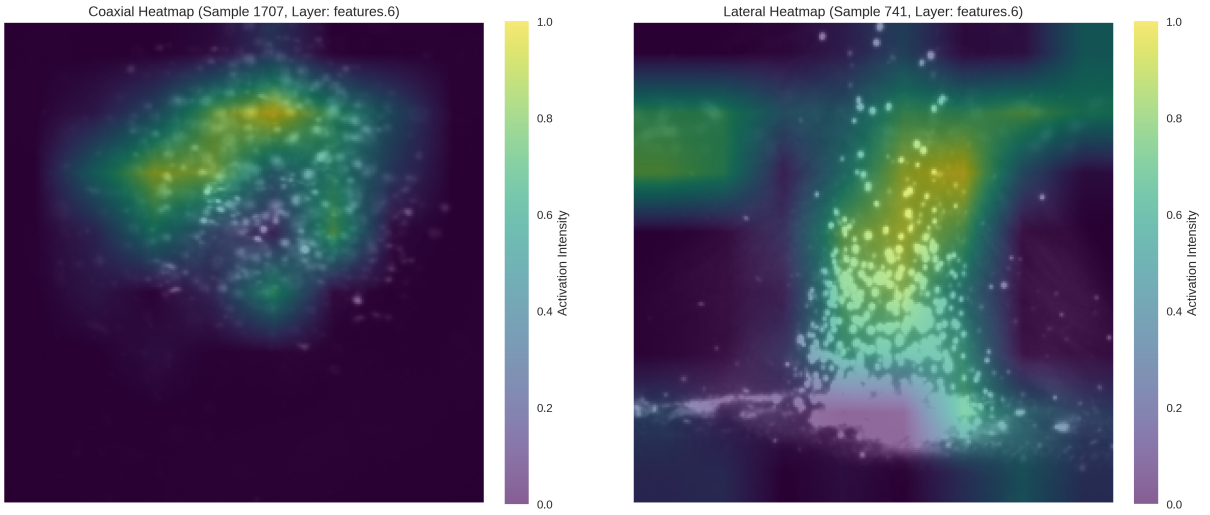
In contrast, the coaxial branch showed a more nuanced progression of attention across layers. Interestingly, the deepest layer features.8 (Figure 41) showed unexpected high activation in the top-right corner, where no process-related activity is visible. This could indicate either noise or a spurious feature that was overemphasized by the network at that depth. However, layers features.7 (Figure 42) and features.6 (Figure 43) demonstrated much more relevant activations, clearly focusing on the melt pool region and symmetry of the deposition area, suggesting that important features were encoded at these intermediate depths.

Figure 42 – Grad-Cam Heatmap - Feature 7



Source: The Author.

Figure 43 – Grad-Cam Heatmap - Feature 6



Source: The Author.

The analysis reveals that while the model successfully learns meaningful visual patterns in both views, not all layers contribute equally to the final prediction. High-level layers may sometimes lose focus or overfit to less relevant spatial cues, emphasizing the importance of visual inspection when evaluating model interpretability.

By combining both views, the model benefits from complementary spatial information: side images capture vertical structure, plume geometry, and material flow, while coaxial images contribute information about melt pool symmetry, distribution of bright spots, and intensity gradients. These interpretations help justify the design choice of a dual-encoder architecture and strengthen confidence in the model's reliability.

Although Grad-CAM maps were also generated for earlier convolutional layers (features.0 to features.5), they were excluded from the main discussion due to their low-level and less interpretable nature. These layers primarily activate in response to basic visual patterns such as edges, textures, and local contrast, which are essential for feature extraction but do not provide clear semantic indications of the model's attention. The full set of visualizations is included in Appendix A for completeness and transparency.

To enhance the reliability of Grad-CAM interpretations, a TabNet combiner was used during feature fusion instead of a simple concatenation layer. The TabNet mechanism applies attentive feature selection at each decision step, allowing the model to dynamically focus on the most informative latent representations from the coaxial and side image branches. In practice, this modification filtered out less relevant activations, such as those found in deeper layers pointing to background areas, and improved the semantic clarity of Grad-CAM results, all while preserving the original prediction accuracy. This strategy provided stronger assurance that the model was relying on

physically meaningful features rather than noise or spurious patterns.

4.6 Final Considerations

The experiments presented in this chapter revealed the CNN model's ability to predict disk speed with high accuracy based on image data from the EHLA process. Across multiple configurations and datasets, the model consistently performed well within the trained operating range (4%–16%), achieving low prediction errors and fast inference times suitable for real-time applications.

Side-view images were identified as the most informative input modality, while coaxial images and carrier gas flow added robustness and interpretability. The best model achieved an MAE ~0.12% of disk speed and R^2 of 0.9980, supported by Grad-CAM visualizations highlighting relevant regions such as the melt pool and powder plume.

Models trained on datasets with removed intermediate steps demonstrated limited interpolation capabilities, with predictions tending to snap to the closest trained values. Although the model was trained using a regression objective, the results revealed a behavior typically associated with classification: the predictions tended to cluster around the known training targets, forming discrete bands rather than a smooth distribution.

Despite these nuances, the models remained reliable for detecting shifts in powder feeding behavior, with response time and error margins aligned with the needs of process monitoring. The findings presented in this chapter form a solid foundation for advancing toward interpretable, image-based monitoring tools in EHLA.

5 Conclusion and Future Work

5.1 Conclusion

This study presented the development and evaluation of a deep learning-based system for predicting disk speed in EHLA using synchronized coaxial and side-view images. The proposed approach demonstrated strong performance within the trained ranges, achieving low prediction errors (MAE ~0.12% of disk speed) and high correlation (R^2 ~0.998), while maintaining fast inference times suitable for practical integration. Grad-CAM analyses confirmed that the model focused on physically meaningful regions of the process, powder stream, melt pool, and the plume above the interaction zone, indicating that it successfully learned relevant features rather than relying on background noise or artifacts.

Importantly, the system proved effective in scenarios aligned with its intended application: detecting deviations in powder feed, such as those caused by clogging events or system misalignments. The model's ability to estimate disk speed accurately from image data provides a valuable non-invasive diagnostic tool for identifying such irregularities, potentially enabling real-time operator feedback and process supervision.

An additional investigation was conducted to assess the influence of different input combinations on model performance. It was observed that the side image alone yielded the highest prediction accuracy (MAE ~0.06% of disk speed), highlighting its critical role in capturing relevant process dynamics. Models that combined side and coaxial images, along with carrier gas flow, achieved balanced results and offered slightly enhanced robustness. In contrast, the model trained with carrier gas flow as the only input performed poorly, reinforcing the dominance of visual information in this task. These findings confirm a trade-off between model complexity, inference time, and accuracy, and suggest that multimodal inputs can enhance interpretability and stability under varied conditions, but may not always lead to superior standalone performance.

During the experimental process, certain limitations were identified, particularly regarding the acquisition of intermediate disk speed samples (Week 3), which were collected under slightly different setup conditions than the initial dataset (Week 2). Despite best efforts to replicate the environment, minor shifts in camera alignment, lighting, and focus introduced domain differences that affected the model's generalization. This was especially evident in interpolation tests, where predictions for excluded disk speeds tended to align with neighboring classes from the same dataset subset, rather than the global scale. Nevertheless, these effects do not invalidate the study. On the contrary, they emphasize the system's sensitivity to domain shifts and underscore the importance of data consistency in vision-based regression tasks.

Overall, the results demonstrate that the proposed solution is both techni-

cally viable and functionally relevant. It lays a solid foundation for future improvements, including enhanced data harmonization, label smoothing for better interpolation, and architectural exploration with vision transformers or alternative encoders. Moreover, the system's modular structure supports its integration into real-time monitoring pipelines, enabling further advancements toward adaptive manufacturing and intelligent diagnostics in laser-based additive processes.

5.2 Future Work

Based on the findings and limitations of this study, several directions for future work are proposed:

1. Label Smoothing and Intermediate Target Generation

   To overcome the limitations of discretely distributed target values, future models could benefit from a smoother label space, for example, by using disk speed values with finer granularity (e.g., 5.1%, 5.2%, ..., 14.9%). Combined with synthetic data generation techniques, such as mixing pairs of images from neighboring classes, this could help introduce intermediate visual patterns and improve interpolation.

2. Robustness through Mask Variation

   To further evaluate model robustness to background noise, future work could explore the application of masks with varied shapes, sizes, and levels of transparency. This approach would test the model's ability to focus on physically relevant features while ignoring irrelevant background information. Additional strategies, such as random background removal or adversarial occlusion, could provide deeper insights into the model's generalization and resilience to visual noise.

3. Architectural Exploration

   In addition to the architectures used in this study, alternative encoders such as ResNet, ConvNeXt, and transformer-based models like ViT (Vision Transformer) or Swin Transformer could be evaluated for both performance and interpretability. Their capacity to capture long-range dependencies might enhance prediction accuracy and stability under degraded visual input.

4. Real-Time Integration

   A key application perspective is the deployment of the trained model in a real-time monitoring system. Two feasible strategies are under consideration:

   - Low-latency UDP inference pipeline, where images are extracted directly from the acquisition buffer and sent to a Linux-based inference module.

- Buffered folder-based streaming, where predictions are made with a short acceptable delay (e.g., 1 second) by processing images stored in a shared directory.

These integrations would allow the model to be tested under realistic conditions and pave the way for real-time operator feedback and process monitoring.

BIBLIOGRAPHY

ASTM. *ASTM F2792 – Standard Terminology for Additive Manufacturing Technologies*. [S.l.]: West Conshohocken, PA: ASTM International, 2012. Available at: https://www.astm.org/. Accessed on: November 2024. 20

BANFI, M. *et al.* Ground control: an acquisition and control system architecture for lmd. *Procedia CIRP*, v. 107, p. 605–610, 2022. Available at: https://www.sciencedirect.com/science/article/pii/S2212827122003171. Accessed on: May 2025. 24

BASLER. *Basler Machine Vision Expert | Basler AG*. 2025. Available at: https://www.baslerweb.com/de-de/. Accessed on: May 2025. 84

BAUER, M. *et al.* Multi-modal artificial intelligence in additive manufacturing: Combining thermal and camera images for 3d-print quality monitoring. SciTePress, Setúbal, Portugal, p. 281–288, 2023. Available at: https://www.scitepress.org/Papers/2023/119675/119675.pdf. Accessed on: Apr 2025. 32

BISHOP, C. M. *Pattern Recognition and Machine Learning*. [S.l.]: Springer, 2006. Available at: https://archive.org/details/patternrecogniti0000bish. Accessed on: November 2024. 25

BONIN, C. *et al.* Leveraging machine learning for predicting and monitoring clogging in laser cladding processes: An exploration of neural sensors. *Journal of Laser Applications*, v. 35, n. 4, p. 042022, 2023. Available at: https://doi.org/10.2351/7.0001154. Accessed on: November, 2024. 16, 17, 22

CHATTOPADHYAY, A. e. a. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. *arXiv preprint arXiv:1710.11063*, 2018. Available at: https://arxiv.org/abs/1710.11063. Accessed on: June 30, 2025. 32

CHEN, L. *et al.* In-situ crack and keyhole pore detection in laser directed energy deposition through acoustic signal and deep learning. *Additive Manufacturing*, v. 69, p. 103547, 2023. Available at: https://www.sciencedirect.com/science/article/pii/S2214860423001604. Accessed on: May 2025. 24

DASS, A.; MORIDI, A. State of the art in directed energy deposition: From additive manufacturing to materials design. *Coatings*, v. 9, n. 7, p. 418, 2019. Available at: https://doi.org/10.3390/coatings9070418. Accessed on: November 2024. 20

DONG, F. *et al.* Cross-section geometry prediction for laser metal deposition layer-based on multi-mode convolutional neural network and multi-sensor data fusion. *Journal of Manufacturing Processes*, v. 108, p. 791–803, 2023. Available at: https://doi.org/10.1016/j.jmapro.2023.11.036. Accessed on: May 2025. 23

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. Available at: https://www.deeplearningbook.org. Accessed on: November 2024. 24, 25, 26, 27, 28, 29

HORNET. *Standard Systems | Hornet Laser Cladding*. 2025. Available at: https://www.hornetlasercladding.com/standard-systems. Accessed on: May 2025. 34

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, v. 25, 2012. Available at: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf. Accessed on: November 2024. 29

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, p. 436–444, 2015. Available at: https://www.researchgate.net/publication/277411157_Deep_Learning. Accessed on: November 2024. 29

LECUN, Y. *et al.* Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. Available at: https://doi.org/10.1109/5.726791. Accessed on: November 2024. 30, 31

LUDWIG AI. *Installation Guide*. 2024. Available at: https://ludwig.ai. Accessed on: November 2024. 32

LUDWIG AI. *Ludwig: A declarative machine learning framework*. 2024. Available at: https://ludwig.ai. Accessed on: November 2024. 32

LUMEN LEARNING. *The Electromagnetic Spectrum*. 2025. Available at: https://courses.lumenlearning.com/suny-astronomy/chapter/the-electromagnetic-spectrum/. Accessed on: May 19, 2025. 84

OERLIKON. *Product Data Sheet Metco Twin 150 Powder Feeder*. 2025. Available at: https://www.oerlikon.com/ecoma/files/DSE-0083.10_TWIN_150_EN.pdf?download=true. Accessed on: May 2025. 34

O'SHEA, K.; NASH, R. *An Introduction to Convolutional Neural Networks*. 2015. Available at: https://arxiv.org/abs/1511.08458. Accessed on: November 2024. 25, 26, 29, 30, 31

PERANI, M. *et al.* Track geometry prediction for laser metal deposition based on on-line artificial vision and deep neural networks. *Robotics and Computer-Integrated Manufacturing*, v. 79, p. 102445, 2023. Available at: https://doi.org/10.1016/j.rcim.2022.102445. Accessed on: May 2025. 23

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 65, n. 6, p. 386–408, 1958. Available online: https://www.ling.upenn.edu/courses/cogs501/Rosenblatt1958.pdf. 26

SCHOPPHOVEN, T.; GASSER, A.; BACKES, G. Ehla: Extreme high-speed laser material deposition. *Laser Technik Journal*, Wiley, v. 14, p. 26–29, 2017. ISSN 1613-7728. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/latj.201700020. Accessed on: November 2024. 16, 20, 21, 22

SELVARAJU, R. R. e. a. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. [S.l.]: IEEE, 2017. p. 618–626. Available at: https://arxiv.org/abs/1610.02391. Accessed on: June 30, 2025. 31, 32

TANG, L. *et al.* Variable powder flow rate control in laser metal deposition processes. *Journal of Manufacturing Science and Engineering*, v. 130, p. 95–102,

2008. Available at: https://repositories.lib.utexas.edu/server/api/core/bitstreams/67a7a7c8-7927-47dd-9cf3-b35b197ac980/content. Accessed on: November 2024. 23

TRUMPF. *Image Processing for Cutting and Welding Applications | TRUMPF*. 2025. Available at: https://www.trumpf.com/de_DE/produkte/laser/sensorik/bildverarbeitungschneid-schweissanwendung/. Accessed on: May 2025. 34

ZENíšEK, J. *et al.* Data integration approach for predictive modelling in additive manufacturing. *Procedia Computer Science*, v. 200, p. 1422–1431, 2022. Available at: https://www.sciencedirect.com/science/article/pii/S1877050922003520. Accessed on: May 2025. 23

ANNEX

Table 3 – Week 1 – Variation of Disk Speed and Carrier Gas

Scenario	Number of Claddings	Disk Speed (%)	Carrier Gas Flow (NLPM)
01	6	0, 10, 25, 50, 75, 100	8
02	5	50	2, 4, 8, 12, 16
03	6	0, 25, 50, 75, 2, 4, 8, 12, 16, 100	8
04	6	0, 10, 25, 50, 75, 100	8
05	5	50	2, 4, 8, 12, 16
06	6	0, 25, 50, 75, 2, 4, 8, 12, 16, 100	8
Total Number of Claddings:	72		

Source: The Author.

Table 4 – Week 1 – Fixed Parameters

Laser Power (W)	Travel Speed (mm/rev)	Shielding Gas (L/min)	Standoff distance (mm)	Process Speed (m/min)
3000	0,5	13	15	80

Source: The Author.

Table 5 – Week 2 – Smaller Range (4 -16%) - Variation of Disk Speed and Carrier Gas

Scenario	Number of Claddings	Disk Speed (%)	Carrier Gas Flow (NLPM)
04	5	4, 7, 10, 13, 16	8
05	5	10	4, 6, 8, 10, 12
06	25	4, 7, 10, 13, 16	4, 6, 8, 10, 12
Total Number of Claddings:	35		

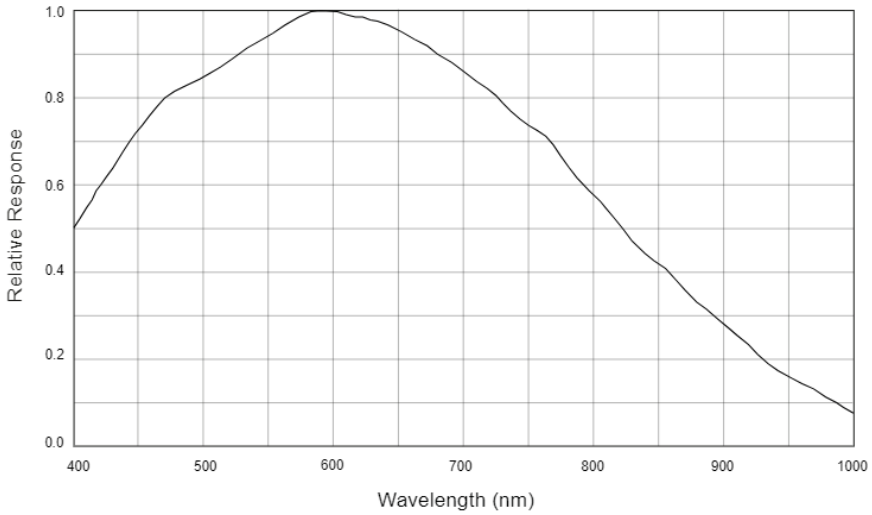
Source: The Author.

Table 6 – Week 2 – Fixed Parameters

Laser Power (W)	Travel Speed (mm/rev)	Shielding Gas (L/min)	Standoff distance (mm)	Process Speed (m/min)
3800	0,1	13	15	80

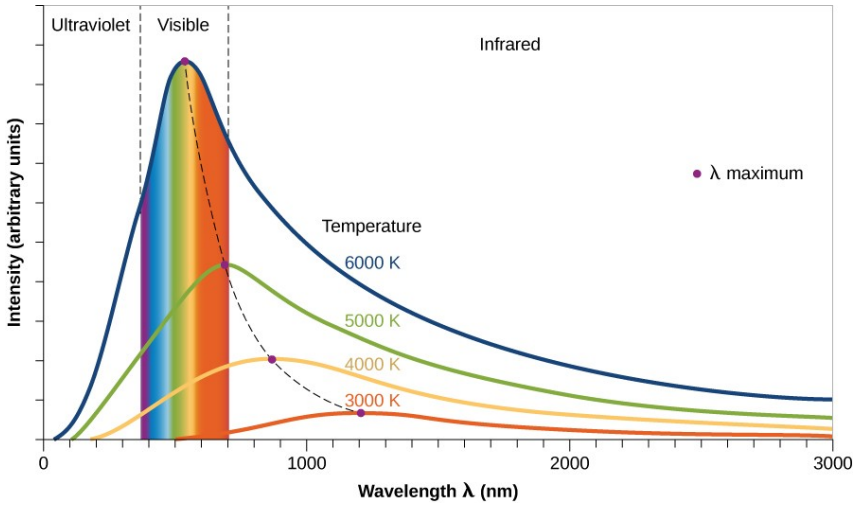
Source: The Author.

Figure 44 – Spectral response curve of the Sony CMOS sensors used in both Basler camera models, highlighting peak sensitivity near 600 nm.



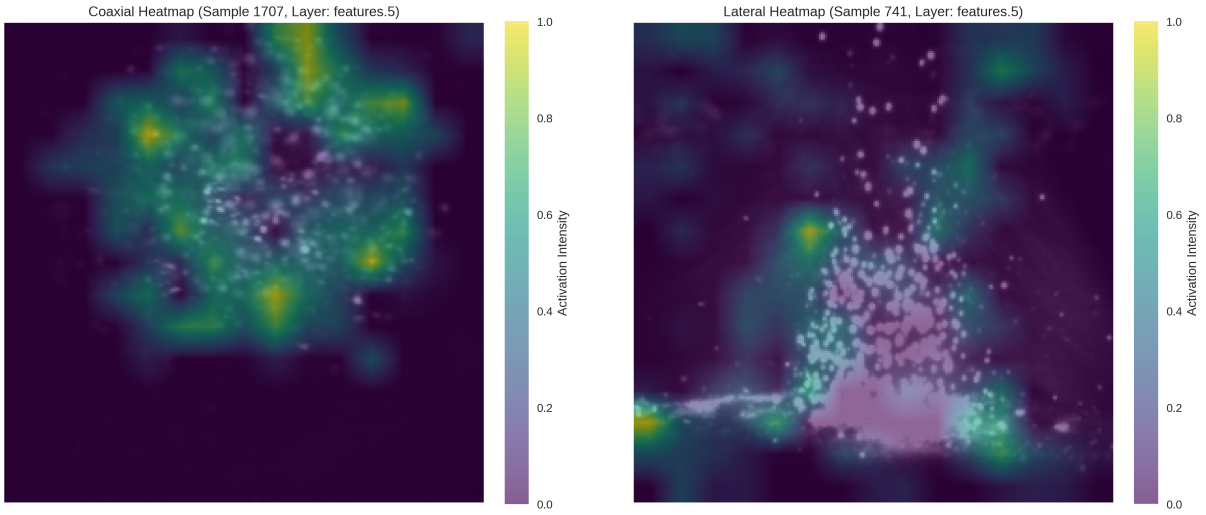
Source: BASLER (2025).

Figure 45 – Illustration of radiation laws



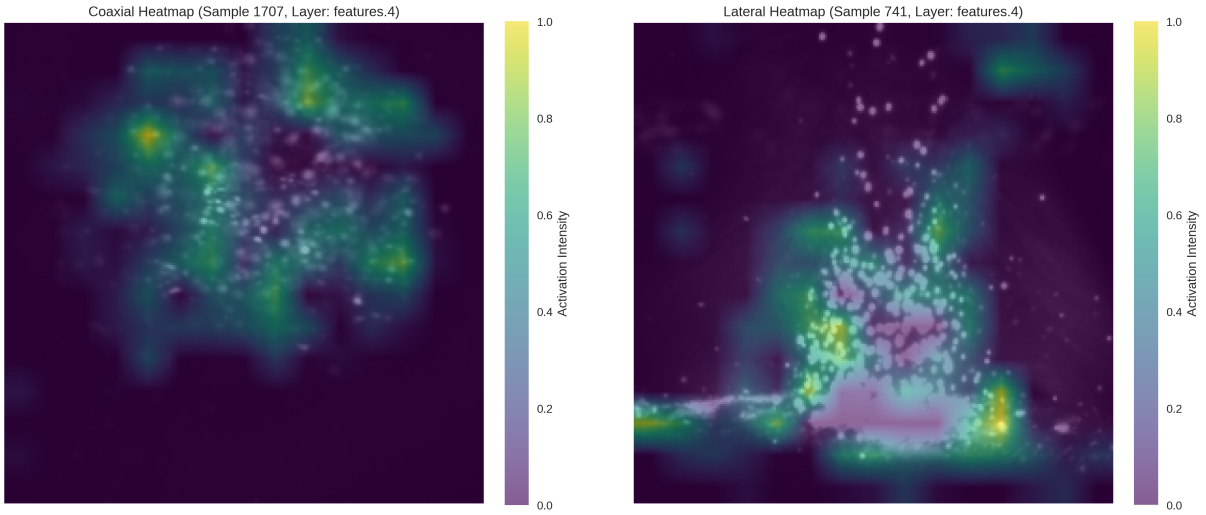
Source: LUMEN LEARNING (2025).

Figure 46 – Grad-Cam Heatmap - Feature 5



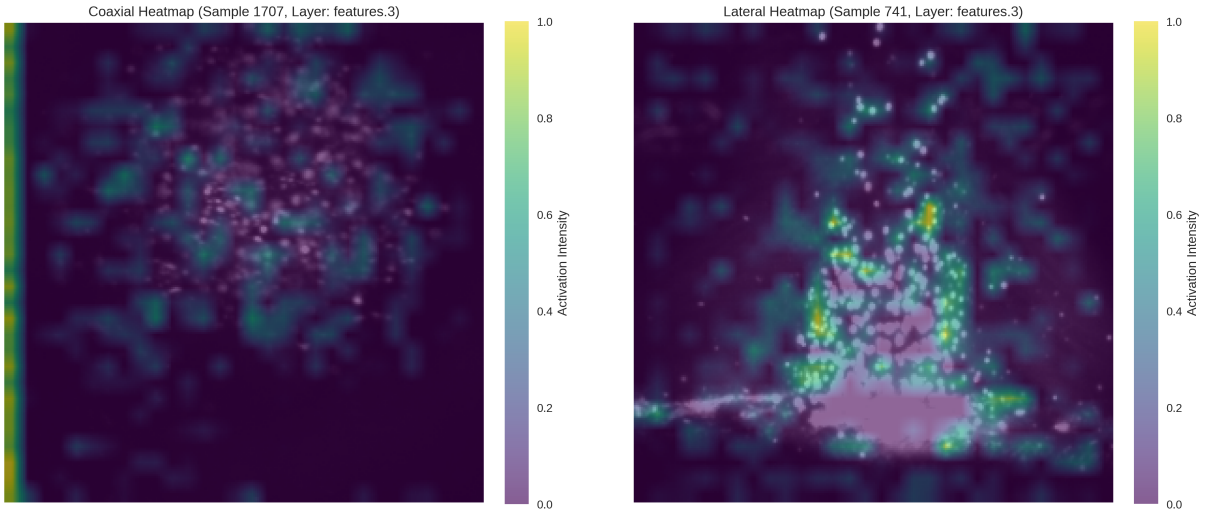
Source: The Author.

Figure 47 – Grad-Cam Heatmap - Feature 4



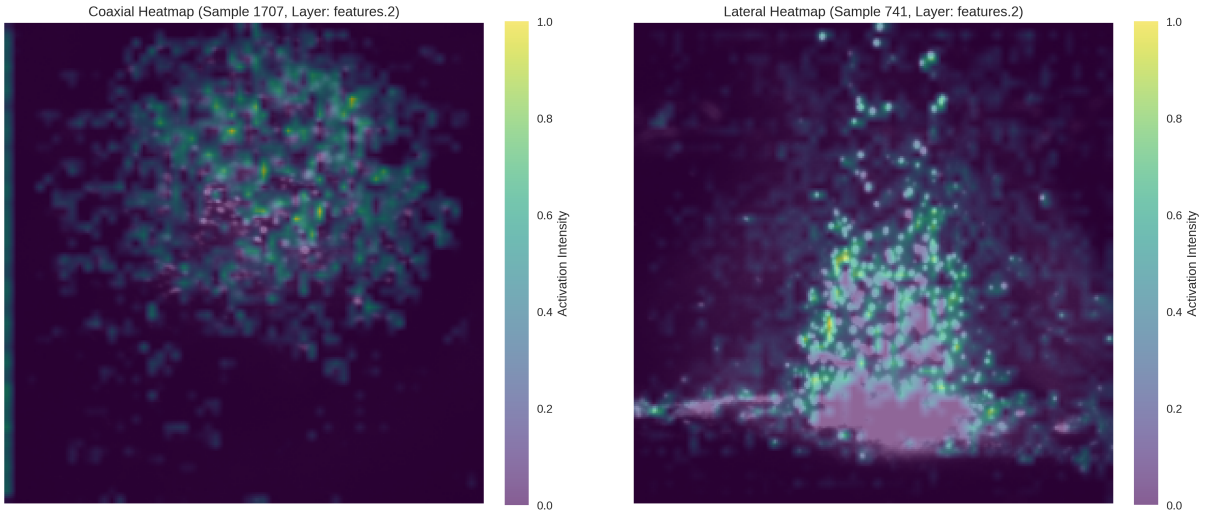
Source: The Author.

Figure 48 – Grad-Cam Heatmap - Feature 3



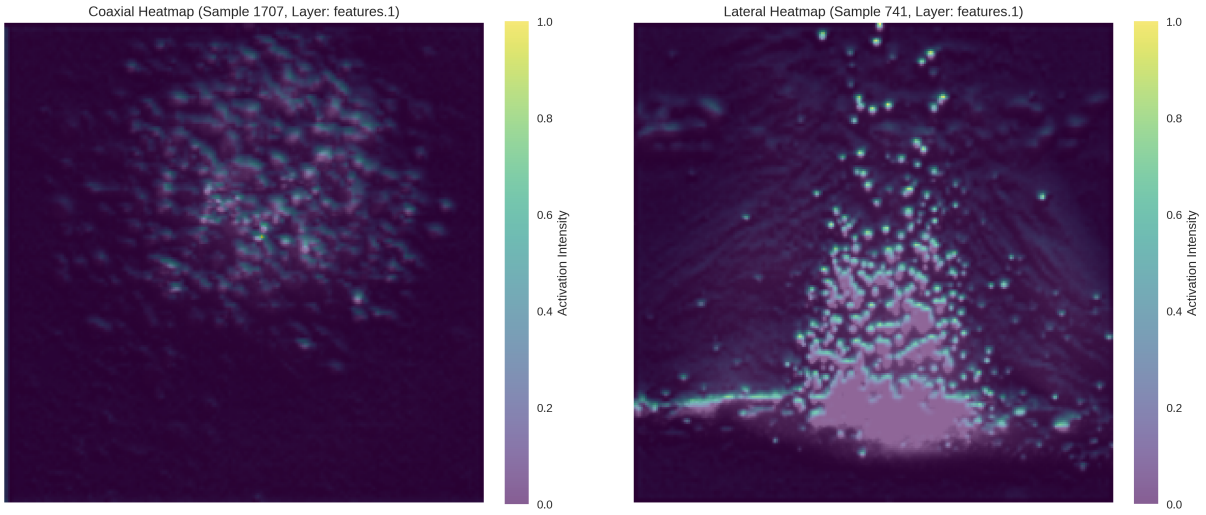
Source: The Author.

Figure 49 – Grad-Cam Heatmap - Feature 2



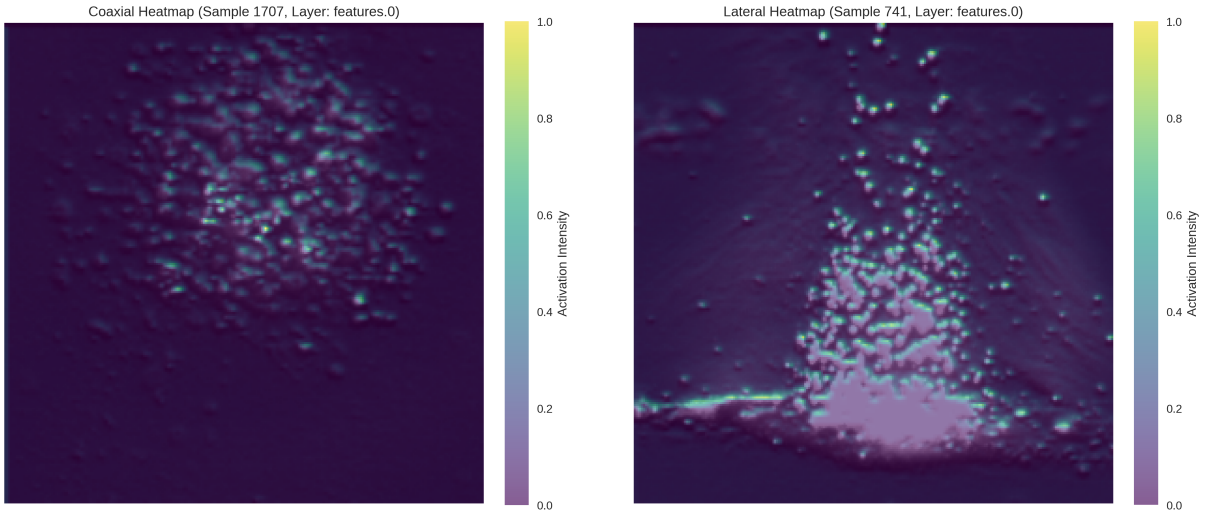
Source: The Author.

Figure 50 – Grad-Cam Heatmap - Feature 1



Source: The Author.

Figure 51 – Grad-Cam Heatmap - Feature 0



Source: The Author.