

Article Investigation of Laser Ablation Quality Based on Data Science and Machine Learning XGBoost Classifier

Chien-Chung Tsai * and Tung-Hon Yiu

Department of Semiconductor and Electro-Optical Technology, Minghsin University of Science and Technology, Hsinchu 30401, Taiwan; tonyyiu30@gmail.com

* Correspondence: cctsai@must.edu.tw

Abstract: This work proposes a matching data science approach for the laser ablation quality, r_{eb} , the study of Si₃N₄ film based on supervised machine learning classifiers in the CMOS-MEMS process. The study demonstrates that there exists an energy threshold, E_{th} , for laser ablation. If the laser energy surpasses this threshold, increasing the interval time will not contribute significantly to the recovery of pulse laser energy. Thus, r_{eb} enhancement is limited. When the energy is greater than 0.258 mJ, there exists a critical value of interval time at which the r_{eb} value is relatively low for each energy level, respectively. In addition, the variation of r_{eb} , Δr_{eb} , is independent of the interval time at the invariant point of energy between 0.32 mJ and 0.36 mJ. Energy and interval time exhibit a Pearson correlation of 0.82 and 0.53 with r_{eb} , respectively. To maintain Δr_{eb} below 0.15, green laser ablation of Si₃N₄ at operating energies of 0.258–0.378 mJ can adopt a baseline interval time of the initial baseline multiplied by $1/\sqrt[3]{2}$. Additionally, for operating energies of 0.288–0.378 mJ during Si₃N₄ laser ablation, Δr_{eb} can be kept below 0.1. With the forced partition methods, namely, the k-means method and percentile method, the XGBoost (v 2.0.3) classifier maintains a competitive accuracy across test sizes of 0.20–0.40, outperforming the machine learning algorithms Random Forest and Logistic Regression, with the highest accuracy of 0.78 at a test size of 0.20.

Keywords: machine learning; XGBoost classifier; laser ablation; CMOS-MEMS; k-means



Citation: Tsai, C.-C.; Yiu, T.-H. Investigation of Laser Ablation Quality Based on Data Science and Machine Learning XGBoost Classifier. *Appl. Sci.* **2024**, *14*, 326. https:// doi.org/10.3390/app14010326

Academic Editor: Soshu Kirihara

Received: 30 November 2023 Revised: 22 December 2023 Accepted: 26 December 2023 Published: 29 December 2023



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

Within the realm of complementary metal oxide semiconductor microelectromechanical systems (CMOS-MEMS) fabrication, following the construction of CMOS circuitry upon the silicon substrate, MEMS structures (e.g., sensors, actuators) are subsequently established atop this circuitry. However, this integration process frequently necessitates modifications to the underlying layers [1], or the formation of intricate features that are potentially unattainable through conventional lithography or etching alone. In such scenarios, laser ablation emerges as a valuable tool. The advantageous nature of laser ablation in post-CMOS-MEMS processing stems from its inherent precision, minimal collateral damage to surrounding regions, and its versatility in addressing a diverse range of materials commonly employed in MEMS fabrication. Nevertheless, meticulous control over laser parameters such as power, wavelength, and pulse duration is paramount to mitigate the potential for undesirable effects, including thermal damage and unintentional material alterations.

The emergence of laser ablation in the 1960s, coinciding with the development of lasers, initially served as a cornerstone for fundamental research in physics, particularly in understanding laser-material interactions. Early applications focused on ablating or removing material from surfaces using high-powered lasers. Subsequent advancements in laser technology propelled the development of diverse laser types, each with unique characteristics and applications. Excimer lasers, for instance, with their short wavelengths and high energies, excel at precise material removal [2]. This technological evolution broadened

the reach of laser ablation into various fields, including microelectronics, semiconductor fabrication, medical device manufacturing [3], and cultural heritage preservation. Within the realm of material processing, laser ablation has become a pivotal technique for micro-machining, surface structuring, thin film deposition, and even micro/nano-scale feature creation with intricate patterns. Ongoing research and development in laser technology continue to refine laser ablation methods. The introduction of new laser sources, such as femtosecond lasers, enables even finer control and minimal thermal damage, unlocking the potential for novel applications and deeper integration across various industries.

Our investigation leverages the TSMC/TSRI D35 common use process, which employs distinct materials for each layer (poly, oxide, via, metal, and silicon nitride (passivation)) to accommodate diverse design configurations. We meticulously fine-tuned a comprehensive array of laser parameters encompassing wavelengths, energy levels, interval time, pulse shots, and pad position, with the explicit objective of precisely removing silicon nitride [4]. The unique properties of Si₃N₄, including its high mechanical strength, thermal stability, chemical inertness, and superior electrical insulating capabilities, have propelled its widespread adoption across various industries. In microelectronics, Si₃N₄ serves as a crucial passivation layer, insulator, or mask material due to its exceptional electrical insulation properties, safeguarding underlying semiconductor devices from external influences. Furthermore, its compatibility with CMOS technology makes Si₃N₄ an extensively utilized structural material in MEMS fabrication, offering both mechanical stability and resilience, while seamlessly integrating with the underlying CMOS circuitry.

Prior to the widespread adoption of laser ablation for Si_3N_4 removal in post-processing of CMOS-MEMS devices, several alternative methods were employed. Chemical etching, utilizing solutions such as hydrofluoric acid (HF), offered selective removal of silicon nitride while minimizing effects on other materials. Dry etching techniques, including reactive ion etching (RIE) and plasma-enhanced chemical vapor deposition (PECVD), were also implemented. These methods relied on plasma reactions for targeted removal of Si_3N_4 . Mechanical approaches, such as polishing and grinding, were occasionally employed, albeit with limitations in their precision. Although chemical and dry etching techniques demonstrated effectiveness, they often struggled to achieve the precise and selective Si_3N_4 removal observed with laser ablation in CMOS-MEMS processing, particularly with regard to minimizing impacts on adjacent materials.

The coalescence of advanced artificial intelligence (AI) with domain-specific engineering expertise is significantly redefining the trajectory of contemporary engineering practice. Recent advancements in temperature monitoring [5], indoor air quality monitoring [6], thermal noise decoupling [7,8], and combustion monitoring [9] exemplify the transformative impact of AI across diverse engineering applications. The strategic integration of field-specific data and features within intelligent models unlocks unprecedented levels of automation and predictability. This synergistic dynamic empowers engineers to optimize resource allocation, enhance design and operational efficiency, automate routine tasks, and unveil novel insights through data-driven discovery.

This study uses the XGBoost algorithm [10] compared with famous classifiers, which are Logistic Regression [11] and Random Forest [12] in supervised machine learning (ML) to categorize the quality of laser ablation on Si_3N_4 film. The study imports the dataset into the unsupervised ML k-means algorithm to label the data by a k value. This intrinsic property of the k value is elaborated by the data science analysis.

XGBoost, Logistic Regression, and Random Forest stand out as widely recognized and highly effective techniques in predictive modeling tasks. Developed by Chen and Guestrin (2016) [10], XGBoost is a gradient boosting algorithm known for its exceptional performance in a range of applications, including regression, classification, and ranking.

XGBoost excels in its capacity to tackle diverse and high-dimensional datasets while mitigating overfitting, a ubiquitous challenge in machine learning. Leveraging an ensemble of decision trees, XGBoost iteratively optimizes a loss function, enabling it to capture complex relationships and achieve state-of-the-art performance. While other algorithms possess distinct strengths, XGBoost offers a compelling combination of versatility and accuracy. Logistic Regression, for instance, presents a clear and interpretable model for binary classification tasks. Its simplicity facilitates understanding the underlying relationships within the data. Random Forest, conversely, capitalizes on an ensemble of decision trees to bolster predictive power and manage overfitting, offering an alternative approach to complex problems.

Due to its applicability in contemporary booming industries such as system on chip (SoC), chip on wafer on substrate (CoWoS) packaging, and CMOS-MEMS chip fabrication, laser micromachining has garnered significant global attention and resources in semiconductor post-processing. The research led by J.A. Grant-Jacob [13,14] at the Optoelectronics Research Centre, University of Southampton, UK, and Yohei Kobayashi [15,16] at the University of Tokyo, Japan, stand at the forefront. Both teams employ deep learning to simulate images of laser ablation, significantly advancing the fundamental models of laser ablation technology. However, the major limitation of deep learning lies in the discernibility of features. Neither of the two teams has been able to propose distinctly identifiable features linking the practical operational parameters of laser ablation to the final ablation quality. In contrast, when using operational parameters as features for machine learning, the quality of laser ablation, whether good or bad, can be confirmed by more precise machine learning algorithms. This approach identifies the crucial parameters in the practical engineering of laser ablation, leading to immediate effects on the advancement of laser ablation practice.

2. Experimental Methodology and Apparatus

2.1. Experimental Equipment and Process

Devices

The objective of this research is to remove the silicon nitride on top of the "dualfunction switch and humidity" chips, patent no. M479441 [17], Taiwan, as shown in Figure 1. Once silicon nitride is removed, the underlying metal will be exposed, fostering humidity sensing.



Figure 1. (a) A 3D schematic of the "dual-function switch and humidity" chip; (b) top view from microscopy.

The efficacy of laser ablation is extensively evaluated through a battery of metrics that assess its performance across various facets. The ablation rate, quantified as material removal per unit time, stands as a central metric for gauging efficiency, is typically measured in volume or depth per pulse or second. Surface quality, pivotal for assessing the post-ablation topography, considers factors like roughness [18], irregularities, and damage, which are evaluated using techniques such as profilometry, atomic force microscopy (AFM), or scanning electron microscopy (SEM). Dimensional accuracy relies on optical or laser systems to ensure precise realization of desired geometries and dimensions. Selectivity evaluates the ability of laser ablation to target material precisely while minimizing collateral damage to adjacent layers, while uniformity guarantees consistent results across the entire substrate. Assessing residual stress and heat-affected zones post-ablation is crucial

for determining potential structural integrity issues. Finally, process efficiency and cost meticulously examine the overall economic viability of laser ablation, encompassing factors like energy consumption, equipment maintenance, and throughput.

The laser ablation quality r_{eb} adopted in this study was originally defined by Tsai and Chan [19,20]. It is a percentage of area ablated within the square laser aperture. Figure 2a shows a sample that underwent laser ablation. The areas of the laser aperture (28.6 × 28.6 µm²) are represented by black squares outlined with solid lines. In Figure 2b, the areas ablated within the square laser aperture are highlighted in blue for r_{eb} calculation. The 532 nm green light laser cutting system, New Wave Research Ezlaze II was utilized, as shown in Figure 3.



Figure 2. A laser ablated sample (**a**) under optical microscope; (**b**) with ablated areas within the square laser aperture delineated.



Figure 3. TSRI laser cutting system–New Wave Research Ezlaze II (New Wave Research, Sunnyvale, CA, USA).

2.2. Methodology

2.2.1. Hypothesis

There exists an energy threshold E_{th} for laser ablation, wherein if the laser energy surpasses this threshold, increasing the interval time will not contribute significantly to the recovery of pulse laser energy. Thus, r_{eb} enhancement is limited.

The Interval Time Baseline (bit) is a defined set of interval times carefully optimized to achieve high laser ablation quality (r_{eb}) with a short interval time for a specific range of energy levels.

There is an invariant point of energy where, under such an energy operation, the variation (Δr_{eb}) of r_{eb} is independent of the interval time.

2.2.2. Methods

In this study, we employed the "dual-function switch and humidity" chips manufactured using the TSMC 0.35 μ m process. The chips underwent a cleaning process involving acetone, isopropyl alcohol, and deionized water. Subsequently, laser shots were fired at the Si₃N₄ passivation of the chips, and the resulting ablated areas within the square laser aperture were then delineated, allowing for the calculation of r_{eb} values.

2.2.3. Study Design and Analysis

This study encompasses two distinct datasets: one comprising 90 samples, and another with 48 samples.

The 90-sample dataset incorporates measurements from five discrete energy levels: 0.258, 0.288, 0.318, 0.348, and 0.378 mJ. For each of these energy levels, we evaluate three distinct interval times: the bit, the bit/ $\sqrt[4]{2}$, and bit/ $\sqrt{2}$. This results in a total of 15 unique combinations. The bit group is established with interval times set at 68, 76, 84, 92, and 100 s. The interval times of bit/ $\sqrt[4]{2}$ group and bit/ $\sqrt{2}$ group are specifically adjusted to $1/\sqrt[4]{2}$ and $1/\sqrt{2}$ times of the bit group, respectively.

The 48-sample dataset comprises data from two energy levels, namely, 0.258 and 0.318 mJ. For each of these energy levels, we evaluate four distinct interval times: the bit, the bit/ $\sqrt[4]{2}$, bit/ $\sqrt{2}$, and bit/2. This results in a total of 8 unique combinations. The bit group is established with interval times set at 68 and 84 s. The interval times of bit/ $\sqrt[4]{2}$ group, bit/ $\sqrt{2}$ group, and bit/2 are specifically adjusted to $1/\sqrt[4]{2}$, $1/\sqrt{2}$, and 1/2 times of the bit group, respectively.

Additionally, for both datasets, three different pulse shots and two distinct pad positions are considered in the data collection process.

2.3. Unsupervised Learning Approach

Three machine learning classification algorithms were utilized: Logistic Regression, Random Forest, and XGBoost. The accuracy is defined as Equation (1).

$$Accuracy = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Samples}}$$
(1)

2.3.1. Logistic Regression

Logistic Regression is a statistical model for binary classification tasks. It employs a sigmoid function to map input features to probabilities. Parameters are learned through gradient descent, minimizing a cost function. Regularization can be applied to prevent overfitting. It is widely used due to its simplicity and interpretability.

2.3.2. Random Forest

Random Forest is an ensemble learning method for both classification and regression tasks. It constructs multiple decision trees during training, each using a subset of the data and features. The final prediction is an average (regression) or majority vote (classification) of individual tree outputs. This technique reduces overfitting and enhances accuracy.

2.3.3. XGBoost

XGBoost, short for eXtreme Gradient Boosting, is an ensemble learning algorithm known for its speed and performance. It builds a sequence of decision trees, each correcting the errors of its predecessor. It employs a gradient boosting framework, optimizing a user-specified loss function. Regularization techniques and parallel processing contribute to its efficiency and effectiveness in various machine learning tasks. The detailed pseudo code of XGBoost is provided in the reference [21].

2.4. Forced Partition Methods

Instead of treating the prediction of r_{eb} as a regression problem, this study categorizes samples into different clusters based on r_{eb} . Two forced partition methods were employed: the percentile method and the k-means method.

2.4.1. Percentile Method

For the percentile method, the desired number of clusters (f) was initially determined, and labels range from 0 to f - 1. Subsequently, samples were divided into f clusters based on their r_{eb} percentiles. Each cluster contained either floor (total number of samples/f) or floor (total number of samples/f) + 1 samples. This method ensured an even distribution of the sample count across each cluster.

2.4.2. K-means Method

For the k-means method, samples underwent k-means unsupervised learning, and the means of energy, interval time, and r_{eb} of each initial cluster were calculated. These three mean values were analyzed to ensure sufficient distinctiveness among the initial clusters. Following this, a forced partition method was introduced. This method established boundaries using the arithmetic mean of the r_{eb} means of two adjacent initial clusters. It is conceivable that initial clusters with closely aligned r_{eb} values might merge into a single initial cluster. Samples falling within the same boundary were subsequently categorized into the same final cluster.

3. Results and Discussion

3.1. Examples

Table 1 lists six samples from three Si₃N₄ pads in this study. Energy, interval time, pulse shots, and pad position were the four parameters adopted, with microscopy photos and laser ablation quality r_{eb} attached. With operating energy of 0.378 mJ and pulse shot of 1 time, regardless of interval time, the ablation results show large portions of silvery areas. This is in contrast with our study conducted in 2022, where Tsai and Chan explored the impact of 532 nm laser energy ranging from 0.138 to 0.318 mJ and pulse shots set at 5, revealing significant areas characterized by a distinct dark coloration [19]. The cause of such disparity is owing to the pulse shot count. Across each pad in Table 1, the discrepancy of r_{eb} is not substantial at all, both from visual inspection and from r_{eb} calculation. This suggests a saturation in r_{eb} .

3.2. Correlation Analysis

The heatmap in Figure 4a, derived from the 90-sample dataset, reveals the relationships among the features: energy, interval time, pulse shot, pad position, and laser ablation quality r_{eb} . Pearson correlation coefficients were employed to quantify the associations between these pairs of variables [22].

A strong positive association was observed between energy and interval time with respect to r_{eb} , as evidenced by their respective correlation coefficients of 0.82 and 0.55. This suggests that higher energy levels and longer interval times are significantly linked to increased r_{eb} . Pulse shot and pad position, however, exhibited relatively modest correlations with r_{eb} , at 0.17 and 0.05, respectively. Furthermore, pulse shot and pad position displayed minimal linear relationships with other features, as indicated by their near-zero correlations.

The heatmap of the 48-sample dataset, Figure 4b, provides a nuanced perspective on the interplay between the features and the laser ablation quality r_{eb} . Pearson correlation coefficients revealed moderate positive associations, with energy and interval time exhibiting correlations of 0.62 and 0.53, respectively, with laser ablation quality r_{eb} .

Comparing the heatmap of Figure 4b to Figure 4a reveals a substantial decrease in the correlation between energy and r_{eb} , dropping from 0.82 to 0.62. Additionally, the association between interval time and r_{eb} exhibited a significant reduction, from 0.55 to 0.53. Furthermore, r_{eb} displays a correlation of 0.032 with pulse shot and 0.036 with pad position, both of which are relatively weak associations. The correlation of r_{eb} with pulse shot experienced a noticeable difference, from 0.17 to 0.032. The correlation of interval time with energy also experienced a noticeable difference, from 0.69 to 0.4. It is noting that pulse shot and pad position continue to display correlations close to zero with other features, indicating little to no linear relationship with them.

Energy	Interval Time (s)					Pulse Shots		PAD-Left		PAD-Right				
0.378	$100 \times \frac{1}{\sqrt{2}} = 70.7$ (belongs to the bit/ $\sqrt{2}$ group)					1		$r_{eb} = 0.83$			$r_{eb} = 0.83$			
0.378	3	$100 \times \frac{1}{\sqrt[4]{2}} = 84.1$ (belongs to the bit/ $\sqrt[4]{2}$ group)					1	an an	$r_{eb} = 0.80$			$\overline{r_{eb}} = 0.80$		
0.378 100 (belongs to the bit group)						1 $r_{cb} = 0.84$			$r_{ch} = 0.88$					
energy (mJ) -	1	0.69	-1.5×10 ⁻¹⁶	5.5×10 ⁻¹⁸	0.82	- 0	0 energy (mj) 8	- 1	0.40	-1.4×10 ⁻¹⁶	7.0×10 ⁻¹⁷	0.62	- 1.0	
interval time (s) -	0.69	1	6.4×10 ⁻¹⁷	1.7×10 ⁻¹⁷	0.55	0.	interval time (s)	- 0.40	1	-5.9×10 ⁻¹⁸	-6.3×10 ⁻¹⁷	0.53	0.0	
pulse_shot -	-1.5×10-16	6.4×10 ⁻¹⁷	1		0.17	- 0.	6 pulse_shot	1.4×10 ⁻¹⁶	⁵ -5.9×10 ⁻¹⁸	1	-5.1×10 ⁻¹⁷	0.03	- 0.6	
PAD (L/R) -	5.5×10 ⁻¹⁸	1.7×10 ⁻¹⁷		1	0.05	- 0.	4 PAD (L/R)	- 7.0×10 ⁻¹⁷	-6.3×10 ⁻¹⁷	-5.1×10 ⁻¹⁷	1	0.04	- 0.4	
reb -	0.82	0.55	0.17	0.05	1	- 0.	z	- 0.62	0.53	0.03	0.04	1	- 0.2	
	energy (mJ) -	interval_time (s) -	pulse_shot -	PAD (L/R) -	- qa	— - 0.	0	energy (mJ) -	interval_time (s) -	pulse_shot -	PAD (L/R) -	- da	- 0.0	
			(a)							(b)				

Table 1. Pad microscopy images and their respective r_{eb} with varied interval time.

Figure 4. Heatmap of (a) 90-sample dataset; (b) 48-sample dataset.

The 90-sample dataset encompasses five distinct energy levels, whereas the 48-sample dataset only contains two. Regarding interval times, the 48-sample dataset exhibits four unique values, while the 90-sample dataset has three. The enhanced range of energy

levels in the 90-sample dataset contributes to a more accurate correlation coefficient of 0.82 between energy and r_{eb} . Likewise, the broader range of interval times in the 48-sample data set fosters a more informative correlation coefficient of 0.53 between interval time and r_{eb} .

3.3. Critical Point of Interval Time

Figure 5a,b depicts the relationship between r_{eb} (on the y-axis) and varying levels of interval time and energy. Figure 5a is derived from a dataset of 90 samples, displaying data for three distinct interval times and five energy levels, resulting in a total of 15 boxes. Within each box, the mean is denoted by a blue diamond. Notably, blue diamonds with the same energy level are connected, forming five distinct lines. Each line connects three means corresponding to the same energy level.



Figure 5. Box plot of (a) 90-sample dataset; (b) 48-sample dataset; (c) 90-sample dataset with bit, $bit/\sqrt[4]{2}$, and $bit/\sqrt{2}$ groups specified. (Note: '^' notation stands for the power of a number).

In contrast, Figure 5b is acquired from a dataset of 48 samples and presents data for two energy levels and four unique interval times, yielding a total of eight boxes. Similarly,

the blue diamond represents the mean within each box. In this figure, two lines connect the means of the same energy level.

In both figures, there is a clear upward trend, indicating that as both interval time and energy level increase, r_{eb} tends to also increase.

To visualize the bit, $bit/\sqrt{2}$, and $bit/\sqrt{2}$ group in Figure 5c, the bit represents a line connecting the rightmost blue diamonds for each of the five energy levels. Similarly, the $bit/\sqrt{2}$ forms another line connecting the five middle blue diamonds for each energy level. Likewise, the $bit/\sqrt{2}$ is a line connecting the five leftmost blue diamonds of the five energy levels.

Upon observing Figure 5a, the mean line for the 0.258 mJ energy level shows an increasing trend with interval time. Specifically, r_{eb} initially rises with interval time, indicating a positive response. However, this effect diminishes as interval time increases, resulting in a decreasing rate of increase in r_{eb} . In contrast, the other energy levels in Figure 5a demonstrate a V-shaped curve. This suggests the presence of an energy threshold E_{th} for laser ablation, likely situated between 0.258 mJ and 0.288 mJ. Beyond this threshold, prolonging the interval time may not significantly contribute to the recovery of pulse laser energy, thereby limiting r_{eb} enhancement. For various laser energies, when the energy exceeds 0.258 mJ, there exists a critical value of interval time (ti)c, at which the r_{eb} value is relatively low. As shown in Table 2, (ti)c is 63.9 s for 0.288 mJ, 70.6 s for 0.318 mJ, 77.4 s for 0.348 mJ, and 84.1 s for 0.378 mJ. The (ti)c were observed in the bit/ $\sqrt[4]{2}$ group.

Table 2. (ti)c of each energy level.

Energy (mJ)	(ti)c (s)
0.288	63.9
0.318	70.6
0.348	77.4
0.378	84.1

In Figure 5b, the mean line for the 0.318 mJ energy level exhibits a pattern similar to that of the 0.258 mJ energy level in Figure 5a. Initially, it rises, but the increase gradually approaches saturation. This is in contrast to the mean line for the 0.318 mJ energy level in Figure 5a, where the line follows a V-shaped curve. In Figure 5b, the mean line for the 0.258 mJ energy level stands out with a sudden spike at an interval time of 57.2 s, creating a distinctive curve not replicated in the other lines of both Figure 5a,b. There is no (ti)c observed from the 48-sample dataset.

Figure 5b reveals a clear similarity between the behavior of the mean line for the 0.318 mJ energy level and that of the 0.258 mJ energy level in Figure 5a. Both exhibit an initial rise, followed by a gradual saturation effect. This contrasts with the V-shaped trajectory observed for the 0.318 mJ energy level in Figure 5a. A defining characteristic of Figure 5b is the prominent spike exhibited by the mean line for the 0.258 mJ energy level at an interval time of 57.2 s. This distinct behavior is not replicated in any other lines across both figures. Notably, the box plot of the 48-sample dataset yielded no (ti)c observations.

3.4. Delta r_{eb}

For each of the 15 combinations of energy levels and interval times (15 boxes) in the dataset of 90 samples, a corresponding range of r_{eb} values (Δr_{eb}) was calculated and is depicted in Figure 6a. The outlier at the leftmost point (an upright triangle) for the energy level 0.258 mJ was excluded in the regression line calculation. Notably, we observe a change in slope ($d\Delta r_{eb}/dt$) from positive to negative in the regression lines.



Figure 6. Δr_{eb} vs. interval time of (a) 90-sample dataset; (b) 48-sample dataset; (c) 90-sample dataset with bit, bit/ $\sqrt[4]{2}$, and bit/ $\sqrt{2}$ groups specified.

Between laser energies of 0.32 mJ and 0.36 mJ there is an invariant point of energy, characterized by a constant Δr_{eb} across all pulse intervals, forming a horizontal line. Under such energy operation, the variation (Δr_{eb}) of r_{eb} is independent of the interval time. Higher energy levels with longer pulse intervals lead to higher r_{eb} values and reduced variability, thereby maintaining high-quality laser ablation.

Conversely, the findings depicted in Figure 6b, which are based on the 48-sample dataset, diverge from those in Figure 6a. Notably, the regression lines for both energy levels, 0.258 mJ and 0.318 mJ, exhibit an identical downward trend in Δr_{eb} as the pulse interval increases.

Considering the variation in Δr_{eb} , to maintain Δr_{eb} below 0.15, green laser ablation of Si₃N₄ at operating energies of 0.258–0.378 mJ can adopt a baseline interval time of the initial baseline multiplied by $1/\sqrt[4]{2}$, as shown in Figure 6c. Additionally, for the operating energies of 0.288–0.378 mJ during Si₃N₄ laser ablation, Δr_{eb} can be kept below 0.1.

The slope values of the regression lines from Figure 6a,b are plotted with energy to generate Figure 7a,b. In Figure 7a, the plot of $d\Delta r_{eb}/dt$ versus energy intersects the x-axis at (0.350 mJ, 0), indicating an energy level at which r_{eb} is independent of interval time.



Figure 7. $d\Delta r_{eb}/dt$ vs. energy of (**a**) 90-sample dataset; (**b**) 48-sample dataset.

In Figure 7b, however, when plotting the two negative slope values and extending the regression line to intersect the x-axis, it yields a negative energy level of -0.197 mJ. This value is inapplicable within the scope of this study.

3.5. Supervised Learning: Classification

At a test size of 0.20, with k-means method k = 5, then f = 4, the accuracies of Logistic Regression and Random Forest are 0.61 and 0.67, respectively. However, using XGBoost

increased the accuracy to 0.78. Therefore, this study adopts XGBoost as the algorithm to analyze the experimental data obtained.

Figure 8a presents the comparative performance of three classification algorithms-Logistic Regression, Random Forest, and XGBoost-in response to diverse test sizes. The algorithms were configured as follows: Logistic Regression was allotted a maximum of 1000 iterations for convergence, Random Forest employed an ensemble of 100 decision trees (n_estimators = 100), and XGBoost utilized the default value of 100 estimators. The x-axis unveils the spectrum of test sizes, spanning from 0.20 to 0.40, while the y-axis illustrates the corresponding accuracy scores. Notably, Logistic Regression consistently exhibits the lowest accuracy among the three models. At a test size of 0.30, a notable crossover occurred where Random Forest surpassed XGBoost, becoming the model with the highest accuracy. The highest accuracy among the three models occurs for the XGBoost algorithm, with an accuracy of 0.74 at a test size of 0.25. For Random Forest, the accuracy varies between 0.56 and 0.67, indicating a consistent performance across different test sizes. XGBoost consistently demonstrates competitive performance, with accuracies ranging from 0.53 to 0.74, showcasing its effectiveness under different test size conditions. XGBoost displays a general decreasing trend in accuracy as the test size increases, suggesting a potential sensitivity to data size. On the other hand, Random Forest maintains a relatively stable trend, indicating robustness to variations in test size.



Figure 8. Accuracies of 90-sample dataset with (**a**) k-means method k = 5, then f = 4; (**b**) percentile method f = 4; (**c**) percentile method f = 5.

Figure 8b illustrates the performance of three classification algorithms using the percentile method. Similar to Figure 8a, Logistic Regression consistently exhibits the lowest accuracy among the three models. This indicates its limited effectiveness in accurately categorizing samples based on the provided features. Notably, at a test size of 0.20, XGBoost demonstrates the highest accuracy with an impressive 0.78. As the test size increases, Logistic Regression shows a gradual decline in accuracy, suggesting potential sensitivity to data size. Random Forest maintains a relatively stable trend in accuracy across the entire range of test sizes, performing between 0.56 and 0.67. This indicates its robustness to variations in test size. Meanwhile, XGBoost starts strong, with a high accuracy of 0.78 at a test size of 0.20. However, it experiences a slight decrease as the test size increases, hinting at a potential sensitivity to data size. Nevertheless, even at the largest test size of 0.40, XGBoost maintains a competitive accuracy of 0.67. XGBoost consistently outperforms Random Forest.

Figure 8c presents the performance of three classification algorithms utilizing a forced partition into five clusters. Consistent with previous observations in Figure 8b, XGBoost consistently outperforms Random Forest across the entire range of test sizes. It is note-worthy that the accuracies obtained in Figure 8c are notably lower compared to those in Figure 8a,b. Despite these lower accuracies, the relative performance of the models remains consistent, underscoring the effectiveness of XGBoost in this dataset. Logistic Regression consistently yields lower accuracy compared to the other models. At a test size of 0.20, XGBoost exhibits the highest accuracy, reaching 0.61. As the test size increases, Logistic Regression shows a gradual decline in accuracy, indicating a potential sensitivity to data

size. Both Random Forest and XGBoost maintain relatively stable trends in accuracy across the range of test sizes. Random Forest performs between 0.41 and 0.52, while XGBoost achieves accuracies ranging from 0.44 to 0.61.

4. Conclusions

This work proposes a matching data science approach for the laser ablation quality, r_{eb} , the study of Si₃N₄ film based on supervised machine learning classifiers in the CMOS-MEMS process. The study demonstrates that there exists an energy threshold, E_{th} , for laser ablation. If the laser energy surpasses this threshold, increasing the interval time will not contribute significantly to the recovery of pulse laser energy. Thus, r_{eb} enhancement is limited. When the energy is greater than 0.258 mJ, there exists a critical value of interval time at which the r_{eb} value is relatively low for each energy level, respectively. In addition, the variation of r_{eb} , Δr_{eb} , is independent of the interval time at the invariant point of energy between 0.32 mJ and 0.36 mJ. Nevertheless, higher energy levels with a longer interval time led to higher r_{eb} values and reduced variation of r_{eb} , thereby maintaining high-quality laser ablation. The results of this research have been validated through data science and machine learning, and they abide by the fundamental principles of laser physics.

5. Future Research Directions

The study of selective laser ablation offers insights not only for Si_3N_4 , but also for another dielectric, SiO_2 , situated at the chip's central white region. The ablation of SiO_2 could alter its physical properties from hydrophobic to hydrophilic. Through capillary phenomena, the air's humidity could penetrate the now hydrophilic SiO_2 surface improving wetting characteristics. This enhancement would allow the CMOS circuit to conduct by the moist sensing functionality of the "dual-function switch and humidity" chip, once voltage-driven.

Author Contributions: Conceptualization, C.-C.T.; methodology, C.-C.T.; formal analysis, C.-C.T. and T.-H.Y.; data curation, T.-H.Y.; writing, C.-C.T. and T.-H.Y.; visualization, T.-H.Y.; supervision, C.-C.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Science and Technology Council, R. O. C., MOST 110-2221-E-159-003

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Acknowledgments: Thanks to the Taiwan Semiconductor Research Institute, TSRI, for providing the equipment. Thanks to Hao-Hsin Gao, Li-Chuan Chen, and Ming-Jie Syu for their implementation of the experiment.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Heinrich, G.; Bähr, M.; Stolberg, K.; Wütherich, T.; Leonhardt, M.; Lawerenz, A. Investigation of ablation mechanisms for selective laser ablation of silicon nitride layers. *Energy Procedia* 2011, *8*, 592–597. [CrossRef]
- Mazalan, M.; Johari, S.; Ng, B.P.; Wahab, Y. Characterization of MEMS Structure on Silicon Wafer using KrF Excimer Laser Micromachining. In Proceedings of the 2014 IEEE International Conference on Semiconductor Electronics (ICSE2014), Kuala Lumpur, Malaysia, 27–29 August 2014.
- Anuar, A.F.M.; Saadon, S.; Wahab, Y.; Fazmir, H.; Zainol, M.Z.; Johari, S.; Mazalan, M. Fabrication of a Cost-Effective MEMS-Based Piezoresistive Cantilever Sensor for Gait Movement Analysis. *Mov. Health Exerc.* 2016, *5*, 25–36.
- Poulain, G.; Blanc, D.; Focsa, A.; De Vita, M.; Semmache, B.; Gauthier, M.; Pellegrin, Y.; Lemitia, M. Laser Ablation Mechanism of Silicon Nitride Layers in A Nanosecond UV Regime. *Energy Procedia* 2012, 27, 516–521. [CrossRef]
- 5. Shin, S.; Ko, B.; So, H. Noncontact thermal mapping method based on local temperature data using deep neural network regression. *Int. J. Heat Mass Transf.* 2022, 183, 122236. [CrossRef]

- 6. Shin, S.; Baek, K.; So, H. Rapid monitoring of indoor air quality for efficient HVAC systems using fully convolutional network deep learning model. *Build. Environ.* **2023**, 234, 110191. [CrossRef]
- 7. Baek, K.; Shin, S.; So, H. Decoupling thermal effects in GaN photodetectors for accurate measurement of ultraviolet intensity using deep neural network. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106309. [CrossRef]
- 8. Bang, J.; Baek, K.; Lim, J.; Han, Y.; So, H. Deep Neural Network Regression-Assisted Pressure Sensor for Decoupling Thermal Variations at Different Operating Temperatures. *Adv. Intell. Syst.* **2023**, *5*, 2300186. [CrossRef]
- Shin, S.; Kwon, M.; Kim, S.; So, H. Prediction of Equivalence Ratio in Combustion Flame Using Chemiluminescence Emission and Deep Neural Network. Int. J. Energy Res. 2023, 2023, 3889951. [CrossRef]
- 10. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
- 11. Cox, D.R. The Regression Analysis of Binary Sequences. J. R. Stat. Society. Ser. B 1958, 20, 215–232. [CrossRef]
- 12. Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
- 13. Grant-Jacob, J.A.; Mills, B.; Zervas, M.N. Live imaging of laser machining via plasma deep learning. *Opt. Express* **2023**, *31*, 42581–42594. [CrossRef] [PubMed]
- Grant-Jacob, J.A.; Mills, B.; Zervas, M.N. Acoustic and plasma sensing of laser ablation via deep learning. *Opt. Express* 2023, *31*, 28413–28422. [CrossRef] [PubMed]
- 15. Shimahara, K.; Tani, S.; Sakurai, H.; Kobayashi, Y. A deep learning-based predictive simulator for the optimization of ultrashort pulse laser drilling. *Commun. Eng.* 2023, 2, 1. [CrossRef]
- 16. Tani, S.; Kobayashi, Y. Ultrafast laser ablation simulator using deep neural networks. Sci. Rep. 2022, 12, 5837. [CrossRef] [PubMed]
- 17. Tsai, C.-C.; Lin, C.-H. Chip Level Microelectromechanical Humidity Switch. R. O. C. Patent No. M479441, 1 June 2014.
- Wang, L.; Suess-Wolf, R.; Ulm, M.; Franke, J. Investigation of the laser ablation threshold for optimizing Laser Direct Structuring in the 3D-MID Technology. In Proceedings of the 2018 International Conference on Electronics Packaging and iMAPS All Asia Conference (ICEP-IAAC), Mie, Japan, 17–21 April 2018.
- Tsai, C.-C.; Chan, C.-C. Ensemble and Unsupervised Machine Learning Applied on Laser Ablation Quality Study of Silicon Nitride during CMOS-MEMS Post Processing. In Proceedings of the 2022 IEEE 4th Eurasia Conference on IOT, Communication and Engineering (ECICE), Yunlin, Taiwan, 28–30 October 2022.
- Tsai, C.-C.; Chan, C.-C. Laser Ablation Quality Study of Silicon Nitride during CMOS-MEMS post Processing by Using Machine Learning and Data Science. In Proceedings of the 3rd IEEE Eurasia Conference on IOT, Communication and Engineering, Yunlin, Taiwan, 29–31 October 2021.
- XGBoost: The Definitive Guide (Part 1). Medium. Available online: https://towardsdatascience.com/xgboost-the-definitiveguide-part-1-cc24d2dcd87a (accessed on 21 December 2023).
- Pearson Correlation Coefficient. Wikipedia. Available online: https://en.wikipedia.org/wiki/Pearson_correlation_coefficient (accessed on 30 November 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.