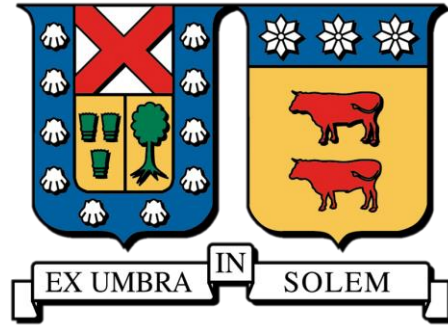




UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS

**Modelo Predictivo para la Detección Temprana de Fallas en Procesos de
Mecanizado CNC mediante Machine Learning.**

TESIS PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

AUTOR

Anaís Millaray Salas Cabello

PROFESOR GUÍA

David Godoy Ramos

PROFESOR CO-REFERENTE

Gonzalo Améstica Hernández

VALPARAÍSO DE CHILE, Junio 2026



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: _Modelo Predictivo para la Detección Temprana de Fallas en Procesos de Mecanizado CNC mediante Machine Learning._

Nombre del candidato(a): _Anaís Millaray Salas Cabello_

Carrera / Grado: _Ingeniería Civil industrial_

Campus: _Vitacura_ Departamento: _Industrias_

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, _Da_ v _id Godoy R., en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (embargo) por (marcar una opción):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: _31-07-2026_ Firma: _____

Estudiante o Candidato(a):

Fecha: _09-10-2026_ Firma: _____

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Tabla de Contenidos

1. Abstract	4
2. Problema de Investigación	5
2.1. Antecedentes y contexto.....	5
2.2. Descripción del problema de investigación	6
2.3. Relevancia del problema	7
2.4. Propuesta de abordaje y alcance	8
3. Objetivos	10
3.1. Objetivo General.....	10
3.2. Objetivos Específicos.....	10
4. Marco Teórico	11
4.1. Gestión de Activos y Modelos Predictivos.....	12
4.1.1. Gestión de Activos Físicos e ISO 55000.....	12
4.1.2. Evolución hacia el Mantenimiento Predictivo 4.0	13
4.2. Ciencia de Datos y Probabilidad	15
4.2.1. Distribución Binomial.....	15
4.2.2. Distribución de Poisson	16
4.2.3. Distribución Normal	17
4.2.4. Modelos Híbridos y Técnicas de Machine Learning	19
4.2.5. Clasificación en Aprendizaje Supervisado	20
4.2.6. Evaluación de Modelos Predictivos y Matrices de Confusión.....	22

4.2.7.	Integración de la simulación y modelos de clasificación en la predicción de defectos.....	24
4.3.	Contexto Industrial y Maquinaria	26
4.3.1.	Evolución de la Manufactura: Hacia la Industria 4.0.....	26
4.3.2.	Integración Tecnológica: IoT y Computación en la Nube.....	27
4.3.3.	Maquinaria CNC y Variables Operacionales Críticas	28
4.4.	Síntesis del Marco Teórico y Vinculación con el Problema	30
5.	Metodología	31
5.1.	Fase 1: Modelación Estocástica del Entorno CNC	31
5.1.1.	Modelación de Variables Operacionales Continuas (Distribución Normal) .	32
5.1.2.	Severidad y Conteo de Anomalías (Distribución de Poisson)	35
5.1.3.	Etiquetado Binario del Estado del Sistema (Distribución Binomial).....	37
5.1.4.	Parametrización y Justificación de Variables Físicas	39
5.2.	Fase 2: Generación, Procesamiento y Estructuración del Conjunto de Datos	41
5.2.1.	Extracción y Consolidación Matricial de Señales	42
5.2.2.	Ingeniería de Características y Prevención de Fuga de Información	43
5.2.3.	Estructuración Algorítmica de la Probabilidad de Falla	44
5.3.	Fase 3: Entrenamiento y Evaluación de Algoritmos Predictivos	45
5.3.1.	Implementación en Entorno Classification Learner	46
5.3.2.	Estrategia de Validación Cruzada	47
5.3.3.	Criterios de Selección: Matriz de Confusión y Costo de la No Calidad . . .	48
6.	Resultados	51

6.1. Análisis de Sensibilidad del Horizonte de Simulación.....	51
6.1.1. Escenario Base: Simulación de 10 Segundos.....	52
6.1.2. Escenario Intermedio: Simulación de 100 Segundos.....	54
6.1.3. Escenario Óptimo: Simulación de 1000 Segundos.....	55
6.2. Evaluación de Modelos bajo Aprendizaje Sensible al Costo.....	57
6.2.1. Desempeño de la Red Neuronal Artificial.....	59
6.2.2. Evaluación de Modelos Secundarios	60
6.3. Análisis.....	62
7. Conclusiones	64
7.1. Discusión de resultados.....	64
7.2. Impacto estratégico y operacional.....	66
7.3. Limitaciones y trabajos futuros.....	67

1. Abstract

En la Industria 4.0, los procesos de mecanizado CNC requieren transitar de controles de calidad reactivos a enfoques preventivos debido al impacto de los defectos en la eficiencia y los costos. Dado que los métodos de inspección convencionales presentan limitaciones para la detección temprana de fallas, se evidencia la necesidad de contar con herramientas que permitan anticipar la ocurrencia de defectos y cuantificar el riesgo operativo. La inspección posterior al proceso resulta insuficiente para capturar la relación entre la variabilidad de los parámetros de operación y la severidad de las fallas, lo que se traduce en costos elevados por reprocesos y desperdicios de material. Para abordar esta problemática, se desarrolló una metodología enfocada en el diseño de un modelo probabilístico integrado, basado en distribuciones estadísticas clásicas y simulación estocástica. El objetivo central de la investigación consistió en anticipar la tasa de defectos en procesos CNC mediante la integración de variables operativas críticas, con el propósito de optimizar la calidad y reducir los costos de reproceso. La validación teórica del modelo evidenció su capacidad para identificar de manera adecuada los estados de falla y estimar la magnitud de los defectos con un nivel de precisión aceptable. En consecuencia, se establece que los modelos probabilísticos constituyen una alternativa viable y eficaz para fortalecer la gestión preventiva de la calidad en procesos CNC, favoreciendo la reducción de costos de reproceso y el incremento de la confiabilidad operativa. Finalmente, como trabajos futuros se sugiere extender el alcance del modelo a otros sectores industriales e incorporar técnicas de inteligencia artificial en tiempo real, con el fin de mejorar la interpretación de los resultados y fortalecer la confianza en la automatización de la toma de decisiones.

2. Problema de Investigación

2.1. Antecedentes y contexto

La industria manufacturera global se encuentra en un proceso de transformación impulsado por la denominada Industria 4.0, caracterizada por la integración de sistemas ciberfísicos, automatización avanzada y análisis de datos orientados a mejorar la eficiencia operativa y la calidad de los productos. En este escenario, los procesos de mecanizado mediante Control Numérico Computarizado (CNC) continúan consolidándose como una tecnología clave para la fabricación de componentes de alta precisión en la industria metalmecánica.

A pesar del elevado nivel de automatización alcanzado por estos procesos, la variabilidad inherente a los sistemas productivos persiste como una de las principales fuentes de ineficiencia y pérdidas económicas. Fluctuaciones en variables operativas, tales como condiciones térmicas, fuerzas de corte y parámetros de avance, afectan la estabilidad del proceso y comprometen la conformidad del producto final.

Con el objetivo de reducir dicha variabilidad, la ingeniería industrial ha incorporado metodologías de gestión de calidad como Six Sigma, cuyo principio central consiste en disminuir sistemáticamente la dispersión del proceso hasta alcanzar un nivel de desempeño cercano a la perfección, definido como 3,4 defectos por millón de oportunidades. Esta metodología se fundamenta en el control estadístico de procesos y en la toma de decisiones basada en datos. Sin embargo, en numerosos entornos productivos, la aplicación práctica de estos principios se mantiene bajo un enfoque predominantemente reactivo, basado en la inspección posterior al proceso.

La literatura clásica en gestión de calidad señala que los costos asociados a la mala calidad, conoci-

dos como *Cost of Poor Quality* (COPQ), los cuales incluyen reprocesos, desperdicio de material y paradas no planificadas, pueden representar entre un 15 % y un 20 % de los ingresos por ventas en organizaciones que no logran controlar de manera eficiente la variabilidad de sus procesos (Gryna and Juran, 1999). Este escenario evidencia la necesidad de fortalecer enfoques preventivos en la gestión de la calidad industrial.

2.2. Descripción del problema de investigación

El problema central que aborda esta investigación se relaciona con la limitada capacidad de los enfoques tradicionales de control de calidad para anticipar la ocurrencia de defectos en procesos de mecanizado CNC caracterizados por una alta dinámica operativa. Si bien metodologías como Six Sigma proporcionan un marco robusto para el análisis de causa raíz y la mejora continua, sus herramientas clásicas se apoyan principalmente en el análisis de datos históricos, lo que restringe su capacidad para anticipar fallas durante la ejecución del proceso.

En los procesos de torneado y fresado CNC interactúan múltiples variables operativas de naturaleza estocástica, entre las que destacan la temperatura del husillo, las fuerzas de corte y los parámetros de avance. Las fluctuaciones de estas variables generan deformaciones térmicas y errores dimensionales que afectan directamente la calidad del producto. Estudios recientes evidencian que la optimización de parámetros de mecanizado, como la velocidad de corte y la tasa de avance, resulta fundamental para la sostenibilidad del proceso, aunque difícil de mantener de manera estable en ausencia de sistemas con capacidad predictiva (Cica et al., 2024).

Asimismo, investigaciones más recientes destacan que el monitoreo probabilístico de estas variables permite estimar la calidad del producto antes de su inspección física, aportando información

relevante para la toma de decisiones en entornos de incertidumbre (Papananias, 2025). No obstante, en la práctica industrial, los sistemas de control disponibles carecen de mecanismos que integren de forma efectiva esta información en tiempo real.

El problema se manifiesta cuando los sistemas de control no logran advertir oportunamente que la variabilidad del proceso se aproxima a superar los límites de especificación. Como consecuencia, la detección de defectos ocurre una vez consumido el recurso productivo, generando reprocesos, desperdicio de material y pérdida de eficiencia operativa. Esta situación refleja una brecha entre las capacidades de análisis descriptivo existentes y la necesidad de mecanismos preventivos que permitan anticipar la no conformidad del producto.

2.3. Relevancia del problema

La relevancia del problema abordado en esta investigación se fundamenta en su impacto técnico, económico y metodológico en el ámbito de la manufactura avanzada. Desde una perspectiva técnica, la incapacidad de anticipar defectos limita el aprovechamiento pleno de las tecnologías de automatización y digitalización propias de la Industria 4.0, reduciendo la confiabilidad de los procesos de mecanizado CNC.

Desde el punto de vista económico, la persistencia de enfoques reactivos de control de calidad contribuye al aumento de costos asociados a reprocesos, desperdicio de material y consumo energético innecesario, afectando la competitividad de las organizaciones manufactureras. Esta situación resulta consistente con los elevados valores de COPQ reportados en la literatura clásica de gestión de calidad, los cuales reflejan el impacto financiero de la variabilidad no controlada (Gryna and Juran, 1999).

En términos metodológicos, la investigación adquiere relevancia al proponer un enfoque que integra estadística clásica, simulación estocástica y análisis predictivo como complemento a las herramientas tradicionales de Six Sigma. Este enfoque permite avanzar hacia una gestión preventiva de la calidad, alineada con los principios de mejora continua y toma de decisiones basada en datos, aportando una base conceptual para el desarrollo de sistemas más avanzados de monitoreo y control de procesos industriales.

2.4. Propuesta de abordaje y alcance

Para dar respuesta a la problemática planteada, la investigación propone el desarrollo de un modelo predictivo implementado en un entorno de simulación, orientado a representar de manera formal el comportamiento de un proceso de mecanizado CNC bajo condiciones de variabilidad. La estrategia metodológica se fundamenta en la modelación matemática de la variabilidad del proceso, elemento central de la metodología Six Sigma, mediante el uso de distribuciones estadísticas clásicas y técnicas de simulación estocástica.

A partir de este enfoque, se generan datos sintéticos que reproducen tanto condiciones normales de operación como escenarios de falla, permitiendo analizar el efecto de las desviaciones en las variables críticas sobre la conformidad del producto. El uso de datos simulados posibilita explorar distintos escenarios operativos sin incurrir en los costos y riesgos asociados a la experimentación directa en entornos productivos reales.

Posteriormente, los datos generados se utilizan como base para el entrenamiento y evaluación de algoritmos de inteligencia artificial, con el propósito de estimar la probabilidad de no conformidad del producto antes de su inspección física. Esta integración entre estadística avanzada, simulación

estocástica y gemelos digitales simplificados incorpora una capacidad predictiva al análisis del proceso, complementando los enfoques tradicionales de control de calidad descritos en la literatura reciente (Cica et al., 2024; Papananias, 2025).

El alcance de la investigación se enmarca en un entorno teórico-computacional, por lo que los resultados obtenidos se orientan a demostrar la viabilidad conceptual y metodológica del enfoque propuesto. De este modo, la investigación busca aportar una herramienta analítica que fortalezca la gestión preventiva de la calidad en procesos de manufactura avanzada y sirva como base para futuras aplicaciones en contextos industriales reales.

Bajo este contexto, la presente investigación plantea la siguiente hipótesis de trabajo: La integración de simulación estocástica y algoritmos de Machine Learning entrenados bajo una matriz de costos asimétrica permite anticipar las desviaciones de calidad en el mecanizado CNC, priorizando la detección preventiva de defectos de manera más efectiva que los enfoques estadísticos tradicionales.

3. Objetivos

3.1. Objetivo General

Diseñar un modelo probabilístico híbrido de predicción de defectos en procesos de mecanizado CNC, mediante la integración de simulación estocástica y algoritmos de aprendizaje supervisado, con el fin de anticipar la variabilidad del proceso y reducir los costos operativos asociados a la no calidad.

3.2. Objetivos Específicos

- Determinar y parametrizar las variables operacionales críticas del proceso de mecanizado CNC, temperatura, presión y velocidad, junto con sus distribuciones de probabilidad y rangos de variabilidad, a partir de la revisión de literatura especializada, con el propósito de definir los parámetros de entrada del modelo de simulación estocástica.
- Generar un conjunto de datos sintéticos representativos de la dinámica de fallas y la severidad de defectos del proceso, mediante el diseño de una arquitectura de simulación estocástica.
- Entrenar algoritmos de Inteligencia Artificial utilizando el conjunto de datos generados, para clasificar automáticamente la conformidad de las piezas en función de las desviaciones de los sensores.
- Evaluar la precisión y confiabilidad del modelo propuesto mediante matrices de confusión y métricas de desempeño, analizando su viabilidad como herramienta de apoyo para la reducción de la variabilidad.

4. Marco Teórico

El presente capítulo expone los fundamentos teóricos y conceptuales que sustentan el desarrollo de la investigación. Para abordar la complejidad de los procesos de manufactura avanzada desde un enfoque preventivo, resulta fundamental articular disciplinas de la ingeniería industrial que tradicionalmente se han estudiado de manera aislada.

Con este propósito, el capítulo se estructura en tres ejes principales:

1. **Gestión de Activos e Industria 4.0:** Se contextualiza la evolución de la gestión de activos físicos hacia el mantenimiento predictivo, integrando el paradigma de la digitalización industrial.
2. **Ciencia de Datos y Aprendizaje Supervisado:** Se profundiza en la ciencia de datos, abarcando desde los modelos probabilísticos clásicos que permiten caracterizar la variabilidad de un proceso, hasta los algoritmos de clasificación de *Machine Learning*.
3. **Contexto Operativo CNC:** Se detalla el contexto físico y operativo de la maquinaria CNC, consolidando la base de conocimiento necesaria para justificar las decisiones algorítmicas y de simulación que se implementarán en el marco metodológico.

Esta base teórica permite transitar desde la observación empírica de los procesos industriales hacia su modelación matemática, proporcionando el sustento necesario para el desarrollo del gemelo digital propuesto.

4.1. Gestión de Activos y Modelos Predictivos

La gestión eficiente de los activos físicos es la base para la sostenibilidad operativa. Esta sección describe cómo las estrategias han evolucionado desde el control reactivo hasta la predicción basada en confiabilidad.

4.1.1. Gestión de Activos Físicos e ISO 55000

La gestión de activos físicos se concibe como un enfoque sistemático y coordinado mediante el cual una organización administra sus activos a lo largo de todo su ciclo de vida, con el propósito de generar valor sostenible. Este enfoque abarca desde la planificación y adquisición de los activos, pasando por su operación y mantenimiento, hasta su disposición final, integrando consideraciones técnicas, económicas y de riesgo en la toma de decisiones.

La familia de normas internacionales ISO 55000 establece los principios y lineamientos para implementar un sistema de gestión de activos alineado con los objetivos estratégicos de la organización. De acuerdo con este estándar, la gestión de activos no se limita al control técnico del equipo, sino que implica equilibrar de manera sistemática el desempeño esperado, los costos asociados y los riesgos operacionales. Esta visión integral permite priorizar los recursos disponibles y orientar las decisiones hacia la maximización del valor generado por los activos.

Salazar-Zuluaga et al. (2021) destacan que la adopción de sistemas de gestión de activos basados en ISO 55000 en sectores industriales intensivos en capital, como el energético y el manufacturero, contribuye significativamente a mejorar la coherencia entre la planificación estratégica y la operación diaria. En particular, los autores señalan que este enfoque facilita la identificación de activos críticos y la definición de políticas de mantenimiento más consistentes con el impacto operativo y

financiero de cada equipo.

Asimismo, la literatura resalta que la gestión de activos bajo el marco ISO 55000 promueve el uso estructurado de información técnica y datos operacionales para evaluar el estado y el desempeño de los activos. Esta orientación basada en datos, fuertemente potenciada por la arquitectura de la Industria 4.0, fortalece la capacidad de las organizaciones para anticipar fallas, reducir la incertidumbre y mejorar la confiabilidad de los sistemas productivos Biard and Nour (2021). En procesos de alta precisión, como el mecanizado CNC, esta capacidad resulta especialmente relevante, dado que la condición del activo influye directamente en la estabilidad del proceso y en la calidad del producto final.

En este contexto, la norma ISO 55000 sienta las bases conceptuales para la incorporación de enfoques predictivos en la gestión de activos, al enfatizar la necesidad de comprender el comportamiento futuro de los equipos y su impacto sobre el desempeño global del sistema. Esta perspectiva se alinea con los objetivos de la presente investigación, que busca integrar modelos probabilísticos y herramientas predictivas como soporte a una gestión de activos orientada a la prevención y a la reducción de la variabilidad.

4.1.2. Evolución hacia el Mantenimiento Predictivo 4.0

Las estrategias de mantenimiento industrial han experimentado una evolución significativa como respuesta al aumento de la complejidad de los sistemas productivos y a las crecientes exigencias de confiabilidad, disponibilidad y eficiencia. Tradicionalmente, las organizaciones han operado bajo esquemas de mantenimiento correctivo, en los cuales la intervención se realiza una vez ocurrida la falla, o bajo modelos preventivos basados en intervalos de tiempo o uso. Si bien estos enfoques

permiten reducir la probabilidad de fallas catastróficas, presentan limitaciones importantes al no considerar el estado real del activo ni la variabilidad inherente a su operación.

Hernández-Montero et al. (2023) señalan que la transformación digital impulsada por la Industria 4.0 ha dado lugar al Mantenimiento Predictivo 4.0, el cual integra tecnologías de sensorización, sistemas de adquisición de datos y herramientas avanzadas de análisis para monitorear de manera continua la condición de los activos. Este enfoque permite detectar patrones de degradación y desviaciones operativas antes de que se materialicen en fallas funcionales, favoreciendo una gestión proactiva del mantenimiento. La literatura actual demuestra que esto se logra utilizando técnicas de fusión de sensores para integrar señales de vibración y temperatura en un único modelo predictivo Shuriya (2026), potenciado por algoritmos de aprendizaje automático.

A diferencia del mantenimiento preventivo tradicional, el mantenimiento predictivo se fundamenta en el uso de modelos matemáticos y estadísticos que estiman la evolución futura del estado del activo a partir de variables observables. Toro Lazo (2024) destaca que este enfoque permite optimizar la planificación de las intervenciones, ajustándolas al nivel real de deterioro del equipo, lo que se traduce en una reducción de paradas no planificadas y en un uso más eficiente de los recursos de mantenimiento. Recientes estudios demuestran que el mantenimiento predictivo potenciado con Inteligencia Artificial no solo reduce fallas críticas, sino que disminuye significativamente los incidentes y costos operativos en equipos electromecánicos Loo et al. (2025).

Desde una perspectiva operacional, la implementación del Mantenimiento Predictivo 4.0 impacta positivamente en la Eficiencia General de los Equipos (OEE), al contribuir simultáneamente a mejorar la disponibilidad, estabilizar el rendimiento y preservar la calidad del proceso. En particular,

la anticipación de fallas reduce la variabilidad operativa asociada a detenciones imprevistas y a ajustes correctivos, aspectos críticos en procesos de alta precisión como el mecanizado CNC.

La literatura también destaca que el mantenimiento predictivo no se limita a la prevención de fallas mecánicas, sino que puede extenderse al análisis de desviaciones en variables de proceso que afectan la calidad del producto. En este sentido, el Mantenimiento Predictivo 4.0 se posiciona como un puente entre la gestión de activos y la gestión de la calidad, al proporcionar información temprana que permite intervenir antes de que las desviaciones se traduzcan en defectos.

En el marco de esta investigación, el Mantenimiento Predictivo 4.0 constituye un referente conceptual clave, ya que refuerza la necesidad de incorporar modelos predictivos y probabilísticos capaces de anticipar estados de falla y escenarios de no conformidad. Esta orientación se alinea con el objetivo de reducir la variabilidad del proceso y fortalecer la gestión preventiva de la calidad mediante el uso de simulación y análisis avanzado de datos

4.2. Ciencia de Datos y Probabilidad

Para sustentar la capacidad predictiva del sistema, resulta fundamental definir el marco matemático. Esta sección integra los modelos probabilísticos clásicos orientados al control de procesos con las técnicas modernas de Ciencia de Datos.

4.2.1. Distribución Binomial

La distribución binomial es un modelo fundamental en estadística que se aplica cuando el proceso de producción puede considerarse como una serie de ensayos independientes, en los cuales cada unidad examinada tiene solo dos posibles resultados: éxito (producto sin defecto) o fracaso

(producto defectuoso). Este modelo es especialmente útil en contextos donde el resultado de cada prueba es dicotómico y se conoce la probabilidad de que ocurra un defecto en cada intento. La distribución binomial se caracteriza por su capacidad para predecir la cantidad de defectos esperada en una muestra determinada y para establecer límites de control en los procesos de manufactura.

La naturaleza del experimento se basa en la suposición de que cada unidad o ensayo es independiente, lo que significa que el resultado de un ensayo no influye en los demás. Además, cada ensayo solo puede tener dos resultados posibles: que el producto presente un defecto o que no lo presente.

Este enfoque tiene aplicaciones prácticas muy importantes en el control de calidad. Por ejemplo, permite calcular el número esperado de defectos en un lote de producción, información que es fundamental para diseñar planes de muestreo efectivos y establecer criterios claros de aceptación o rechazo.

Finalmente, su utilidad en el control de calidad se centra en comparar la cantidad real de defectos con la expectativa teórica basada en el modelo. Esta comparación ayuda a identificar desviaciones en el proceso productivo, lo que a su vez permite evaluar la eficacia de los mecanismos de control y detectar problemas de variabilidad antes de que se agraven.

4.2.2. Distribución de Poisson

La distribución de Poisson es otro modelo estadístico clave que se utiliza para describir la ocurrencia de eventos raros a lo largo de un intervalo fijo de tiempo, espacio o cualquier otra dimensión continua. Es especialmente adecuada para situaciones en las que los defectos se producen de manera aleatoria y poco frecuente. A diferencia de la distribución binomial, la distribución de Poisson

se centra en el número de eventos que ocurren en un intervalo dado sin la necesidad de conocer el número total de ensayos, siendo ideal para procesos continuos donde los defectos son esporádicos.

Esta distribución se utiliza cuando los defectos ocurren de forma aleatoria y con baja probabilidad dentro de un intervalo específico. En este contexto, en lugar de requerir el número total de unidades probadas, el modelo se enfoca en la tasa promedio de ocurrencia de defectos, representada por la letra griega λ . Esto es especialmente útil en escenarios en los que los eventos son discretos y poco frecuentes, ya que permite describir la probabilidad de que se produzca un determinado número de defectos sin necesidad de conocer el tamaño total de la muestra.

En el ámbito de la manufactura, la distribución de Poisson permite calcular la probabilidad de observar un número exacto de defectos en un periodo o en un área de producción determinada. Esta característica resulta fundamental para establecer protocolos de intervención. Por ejemplo, si se detecta que la tasa promedio de defectos supera ciertos umbrales críticos, se pueden activar medidas preventivas para evitar que pequeños incrementos en esa tasa se conviertan en problemas mayores. El análisis basado en esta distribución facilita la planificación operativa, ya que permite anticipar desviaciones y ajustar el proceso de producción en tiempo real, contribuyendo así a mantener la calidad y a prevenir un deterioro progresivo de la misma.

4.2.3. Distribución Normal

La distribución normal es un modelo estadístico ampliamente utilizado para describir la variabilidad de variables continuas en procesos productivos, y se conoce popularmente como la curva en campana. Este modelo se fundamenta en el principio de que, cuando se combinan múltiples efectos de pequeños factores independientes, la variable resultante tiende a distribuirse de manera normal

alrededor de una media.

La distribución normal se utiliza para modelar variables continuas que se distribuyen de forma simétrica alrededor de un valor central, caracterizado por la media (μ) y la desviación estándar (σ). Se establecen límites de control, por ejemplo, $\pm 3\sigma$, que ayudan a identificar desviaciones significativas en el proceso.

En manufactura, este modelo es fundamental, ya que permite evaluar la variabilidad en parámetros críticos como dimensiones, tiempos de ciclo y peso de los productos. Al comparar la variabilidad observada con la esperada bajo una distribución normal, se pueden detectar tempranamente posibles problemas en el proceso productivo. Esto facilita intervenciones proactivas con el fin de ajustar la producción y evitar la generación de defectos. En resumen, la distribución normal contribuye a mantener altos estándares de calidad en la fabricación, optimizando la eficiencia operativa y la toma de decisiones en el control estadístico de procesos (SPC), sirviendo como pilar matemático en la aplicación de metodologías de reducción de variabilidad como Six Sigma Garza Ríos et al. (2016). Recientemente, la integración de estos enfoques clásicos de Six Sigma con algoritmos de clasificación como Random Forest ha demostrado elevar sustancialmente la capacidad predictiva del proceso en la industria manufacturera Bratic' et al. (2025).

Para conectar esta variabilidad con el impacto económico, se utiliza la Función de Pérdida de Calidad de Taguchi, la cual postula que cualquier desviación respecto al valor nominal de operación genera una pérdida financiera que crece de forma proporcional al cuadrado de dicha desviación.

Esta filosofía implica que, para el sistema predictivo, reducir la varianza es tan crítico como evitar que el proceso salga de los límites de control. Bajo esta lógica, el modelo de aprendizaje automático

no solo busca clasificar defectos, sino identificar las condiciones operacionales que maximizan la estabilidad económica del proceso.

4.2.4. Modelos Híbridos y Técnicas de Machine Learning

El avance en tecnología de la información ha permitido combinar métodos estadísticos clásicos con técnicas de machine learning, lo que ha dado lugar a la creación de modelos híbridos que ofrecen una mayor precisión en la predicción de defectos. Este enfoque integra lo mejor de ambos mundos, aprovechando las fortalezas de los métodos tradicionales y la capacidad de adaptación y aprendizaje de los algoritmos modernos.

En cuanto a la integración de métodos, un modelo híbrido puede utilizar la distribución binomial para contar la incidencia de defectos y la distribución normal para analizar la variabilidad de parámetros críticos. Estos enfoques estadísticos se complementan con técnicas de machine learning, tales como redes neuronales o árboles de decisión, que son entrenados con grandes volúmenes de datos de producción para identificar patrones complejos que de otra manera serían difíciles de detectar.

Los resultados empíricos han mostrado mejoras notables en la precisión predictiva. Investigaciones recientes, como las desarrolladas por Toro Lazo (2024), destacan que la integración de algoritmos de inteligencia artificial con enfoques probabilísticos incrementa significativamente la capacidad de anticipar fallas en comparación con los métodos tradicionales, permitiendo pasar de un diagnóstico manual a una detección automatizada. Investigaciones recientes han implementado con éxito arquitecturas similares, extrayendo miles de registros desde controladores industriales para alimentar algoritmos de Machine Learning, logrando prevenir paradas de máquina y anticipar productos

defectuosos de forma automatizada Campuzano et al. (2022). Este enfoque permite que el sistema ajuste y optimice automáticamente los parámetros del proceso en función de la retroalimentación continua, lo que se traduce en una mayor eficiencia y calidad en la producción. Investigaciones específicas en maquinaria CNC han validado que los modelos híbridos superan ampliamente a las regresiones estadísticas tradicionales, particularmente en la predicción dinámica de errores térmicos del husillo Lin et al. (2023).

Además, la combinación de estas técnicas ofrece beneficios adicionales al facilitar la identificación de correlaciones no evidentes entre variables operativas y la incidencia de defectos. Esta capacidad analítica apoya la toma de decisiones, especialmente en áreas donde la intervención humana es limitada o demasiado lenta para reaccionar a cambios repentinos en el proceso productivo. En conjunto, la integración de métodos tradicionales y técnicas de machine learning representa un avance significativo en el control de calidad y la optimización de procesos en la manufactura moderna.

4.2.5. Clasificación en Aprendizaje Supervisado

Dentro del aprendizaje supervisado, cuando la variable de salida corresponde a categorías discretas, como Conforme o Defectuoso, el problema se formula como una tarea de clasificación. En el contexto de la manufactura avanzada, este enfoque resulta particularmente relevante, ya que permite transformar variables operacionales continuas, tales como temperatura, presión, velocidad de corte o vibración, en decisiones discretas asociadas a la calidad del producto. De este modo, los algoritmos de clasificación actúan como mecanismos de inferencia que traducen el comportamiento del proceso en una predicción directa sobre la conformidad de la pieza, facilitando la toma de decisiones en tiempo real. Diversos autores han validado que la clasificación supervisada permite

construir relaciones directas entre la telemetría y la calidad final, como la rugosidad superficial, superando los modelos físicos convencionales Guo et al. (2024); Xiao et al. (2024).

Estos modelos se entrenan a partir de datos históricos etiquetados, donde cada observación contiene tanto las condiciones de operación como el resultado final del proceso. A partir de este aprendizaje, el algoritmo es capaz de aproximar una función de mapeo que permite identificar patrones complejos y relaciones no lineales entre variables, superando las limitaciones de los enfoques estadísticos tradicionales cuando la dinámica del sistema es altamente variable o difícil de modelar de forma explícita. En este sentido, la aplicación de técnicas de clasificación se alinea con la necesidad de anticipar defectos antes de la inspección física, contribuyendo a una gestión preventiva de la calidad.

Para abordar este problema, existe una diversidad de enfoques algorítmicos, cada uno basado en distintas lógicas de optimización y segmentación del espacio de características. Entre las técnicas más utilizadas en el análisis de datos industriales destacan:

1. **Árboles de Decisión (Decision Trees):** Corresponden a algoritmos no paramétricos que estructuran el proceso de clasificación mediante una jerarquía de reglas lógicas. En cada nodo, el modelo selecciona la variable y el punto de corte que mejor separa las clases, optimizando criterios como la ganancia de información. Su principal ventaja en entornos productivos radica en su alta interpretabilidad, ya que las decisiones pueden representarse como reglas explícitas, facilitando su validación por parte de los ingenieros de proceso.
2. **Máquinas de Vectores de Soporte (Support Vector Machines - SVM):** El objetivo de una SVM es identificar un hiperplano óptimo en un espacio multidimensional que permita se-

parar las clases maximizando el margen entre dicho hiperplano y los puntos más cercanos, denominados vectores de soporte. En presencia de relaciones no lineales, el algoritmo utiliza funciones kernel que proyectan los datos a espacios de mayor dimensionalidad. Esta propiedad las hace especialmente robustas frente al ruido, característica frecuente en sistemas de medición industrial.

3. **Métodos de Ensamble: Bosques Aleatorios (Random Forest):** Los métodos de ensamble combinan múltiples modelos base, generalmente árboles de decisión, con el objetivo de mejorar la capacidad predictiva y reducir el sobreajuste. Random Forest utiliza técnicas de muestreo aleatorio para generar múltiples árboles independientes, cuya predicción final se obtiene mediante votación. Este enfoque proporciona mayor estabilidad frente a nuevas observaciones, siendo especialmente adecuado para entornos productivos dinámicos.
4. **Redes Neuronales Artificiales (Multilayer Perceptron - MLP):** Inspiradas en estructuras biológicas, las redes neuronales están compuestas por capas de neuronas interconectadas mediante pesos sinápticos y funciones de activación no lineales. Estos modelos aprenden representaciones complejas a partir de los datos mediante procesos iterativos de optimización. Si bien ofrecen una alta capacidad para modelar relaciones no lineales y dinámicas complejas, presentan como principal limitación su menor interpretabilidad, siendo considerados comúnmente como modelos de “caja negra”.

4.2.6. Evaluación de Modelos Predictivos y Matrices de Confusión

El desarrollo de un modelo predictivo no concluye con su entrenamiento; por el contrario, su utilidad práctica depende de una evaluación rigurosa de su desempeño. En problemas de clasificación,

esta evaluación se realiza principalmente a través de la matriz de confusión, herramienta que permite comparar de manera estructurada las predicciones del modelo con los valores reales observados en el proceso productivo.

La matriz de confusión organiza los resultados en cuatro categorías fundamentales: verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. Cada una de estas categorías representa un escenario específico de acierto o error del modelo, siendo particularmente relevante el análisis de los errores en contextos industriales. En el ámbito de la manufactura, los falsos negativos, aquellos casos en los que un producto defectuoso es clasificado como conforme, representan el riesgo más crítico, ya que implican la liberación de productos no conformes hacia el cliente, con potenciales impactos económicos y reputacionales.

A partir de esta estructura, se derivan métricas cuantitativas que permiten evaluar el desempeño del modelo desde distintas perspectivas. La exactitud (accuracy) mide la proporción global de predicciones correctas, proporcionando una visión general del comportamiento del modelo. Sin embargo, en contextos donde las clases están desbalanceadas o donde ciertos errores tienen mayor costo, esta métrica puede resultar insuficiente como único criterio de evaluación.

En este sentido, la precisión (precision) adquiere relevancia al indicar qué proporción de las predicciones de defecto corresponden efectivamente a defectos reales, permitiendo evaluar el impacto de las falsas alarmas en el sistema. Por su parte, la sensibilidad (recall) mide la capacidad del modelo para identificar la totalidad de los defectos presentes, siendo una métrica crítica en sistemas de control de calidad donde la prioridad es minimizar la probabilidad de liberar productos defectuosos.

La consideración conjunta de estas métricas permite establecer un equilibrio entre distintos tipos de error, ajustando el modelo en función de los objetivos operacionales del sistema productivo. En particular, en entornos de manufactura avanzada, este proceso de calibración se orienta a privilegiar la detección temprana de defectos, incluso a costa de un aumento moderado en las falsas alarmas. De esta manera, la evaluación del modelo no solo constituye un ejercicio matemático, sino una herramienta estratégica para la toma de decisiones, alineada con los principios de calidad y mejora continua.

4.2.7. Integración de la simulación y modelos de clasificación en la predicción de defectos

Los elementos teóricos desarrollados en esta sección no se abordan de manera aislada, sino como componentes complementarios dentro de una misma estrategia de modelación orientada a la predicción de defectos en procesos de manufactura.

En primer lugar, las distribuciones de probabilidad, como la binomial, de Poisson y la normal, permiten representar matemáticamente la variabilidad inherente al proceso productivo. Estas herramientas proporcionan la base para la generación de datos sintéticos que replican tanto condiciones normales de operación como escenarios de desviación. En contextos donde los datos históricos de fallas son limitados o incompletos, este enfoque resulta especialmente relevante, ya que permite construir entornos controlados de experimentación sin afectar la operación real.

Sobre esta base, la simulación estocástica actúa como un mecanismo para recrear el comportamiento dinámico del sistema productivo, incorporando incertidumbre en variables críticas del proceso. En la manufactura moderna, a estas representaciones virtuales se les conoce como Gemelos Digitales (Digital Twins), representación virtual de alta fidelidad de un sistema físico, el cual uti-

liza modelos matemáticos y telemetría para simular su comportamiento en tiempo real. Esta tecnología habilitadora permite optimizar operaciones físicas mitigando riesgos operativos Rabanal et al. (2024). La aproximación permite predecir y corregir defectos de maquinado sin arriesgar el equipo físico real Walker et al. (2022); Zhang et al. (2024). El uso de herramientas computacionales para la modelación matemática de estos sistemas complejos, como es el caso del entorno MATLAB/Simulink, permite construir entornos de prueba robustos y seguros Pérez et al. (2022). De este modo, se generan conjuntos de datos que contienen tanto las variables operacionales como la condición final del producto, constituyendo un insumo adecuado para el entrenamiento de modelos de aprendizaje supervisado

Posteriormente, los algoritmos de clasificación analizados en la sección anterior se aplican sobre estos datos con el objetivo de aprender la relación entre las condiciones de operación y la ocurrencia de defectos. Este proceso permite construir modelos capaces de predecir la conformidad de una pieza antes de su inspección física, transformando variables continuas del proceso en decisiones discretas de aceptación o rechazo.

La evaluación de estos modelos, mediante métricas derivadas de la matriz de confusión, permite validar su desempeño y ajustar sus parámetros en función de los objetivos del sistema productivo. En particular, se prioriza la reducción de falsos negativos, dado su impacto crítico en la calidad del producto final.

Dentro de esta investigación, la articulación entre la simulación estocástica y los algoritmos de clasificación representa el fundamento algorítmico esencial, al proporcionar los mecanismos computacionales para modelar la incertidumbre y traducir datos dinámicos en predicciones precisas. Este

enfoque teórico sustenta de manera directa el desarrollo del modelo probabilístico híbrido propuesto para el mecanizado CNC, cuyo propósito es anticipar la variabilidad del sistema y disminuir los costos operativos derivados de la no calidad.

4.3. Contexto Industrial y Maquinaria

El desarrollo de un modelo predictivo requiere no solo de un sólido sustento probabilístico y algorítmico, sino también de una profunda comprensión del entorno físico y tecnológico en el cual operará. Por consiguiente, esta sección aborda el ecosistema de la manufactura avanzada que enmarca la presente investigación.

4.3.1. Evolución de la Manufactura: Hacia la Industria 4.0

La manufactura industrial ha experimentado una evolución progresiva que ha transformado no solo los métodos de producción, sino también la forma en que se gestiona la calidad. Desde la Primera Revolución Industrial, iniciada a mediados del siglo XVIII, se observa una transición desde sistemas artesanales hacia esquemas mecanizados, donde el control de calidad era incipiente y se basaba principalmente en inspecciones visuales posteriores al proceso productivo.

Con la Segunda Revolución Industrial, impulsada por la electrificación, se consolida la producción en masa y la estandarización de procesos. En este período, si bien emergen prácticas más sistemáticas de control de calidad, estas continúan siendo predominantemente reactivas, ya que los defectos se detectan una vez finalizada la fabricación, implicando costos asociados al retrabajo o descarte de productos.

Posteriormente, la Tercera Revolución Industrial introduce la automatización y la informática, per-

mitiendo el desarrollo de herramientas como el Control Estadístico de Procesos (SPC), las cuales representan un avance significativo al incorporar análisis cuantitativo en la supervisión de la calidad. No obstante, estos enfoques aún se centran en el monitoreo del proceso más que en la anticipación de fallas.

En la actualidad, la Industria 4.0 representa un cambio paradigmático al integrar tecnologías digitales avanzadas, tales como inteligencia artificial, Internet de las Cosas (IoT), big data y computación en la nube. Este nuevo enfoque se caracteriza por la implementación de sistemas ciberfísicos que permiten la captura, transmisión y análisis de datos en tiempo real. Como resultado, los sistemas productivos evolucionan hacia modelos predictivos, en los cuales es posible anticipar desviaciones antes de que se traduzcan en defectos, permitiendo una gestión de calidad proactiva y basada en datos. A este cambio de paradigma se le denomina 'Calidad 4.0', la cual busca automatizar la prevención mediante procesos inteligentes que reaccionan en tiempo real para evitar la producción de no conformidades Llanes-Font and Lorenzo-Llanes (2021).

4.3.2. Integración Tecnológica: IoT y Computación en la Nube

La consolidación de la Industria 4.0 se sustenta en la capacidad de interconectar dispositivos físicos mediante el Internet de las Cosas (IoT) y de procesar grandes volúmenes de datos a través de plataformas de computación en la nube. Esta integración tecnológica ha redefinido la forma en que se gestionan los procesos industriales, permitiendo una visibilidad sin precedentes sobre el comportamiento operativo de los sistemas productivos.

A través de sensores instalados en maquinaria y líneas de producción, es posible capturar en tiempo real variables críticas como temperatura, presión, vibración y velocidad de operación. Estos datos,

que anteriormente no eran accesibles o se registraban de forma limitada, constituyen ahora la base para el desarrollo de modelos analíticos avanzados.

La computación en la nube, por su parte, proporciona la infraestructura necesaria para almacenar, procesar y analizar grandes volúmenes de información de manera eficiente y escalable. Esto permite entrenar modelos predictivos con mayor frecuencia y con conjuntos de datos más representativos, mejorando su capacidad de generalización frente a nuevas condiciones operativas.

Asimismo, la integración entre IoT y sistemas analíticos posibilita la implementación de mecanismos de retroalimentación automática. En este contexto, cuando una variable monitoreada supera umbrales críticos previamente definidos, el sistema puede generar alertas en tiempo real o incluso activar respuestas automatizadas. Este tipo de capacidades resulta fundamental para la detección temprana de anomalías, reduciendo la propagación de defectos y fortaleciendo el control del proceso productivo.

4.3.3. Maquinaria CNC y Variables Operacionales Críticas

La aplicación de modelos predictivos requiere su contextualización en un entorno físico específico. La integración de la tecnología CNC en el contexto de la Industria 4.0 ha permitido vincular estas máquinas con sistemas de telemetría IoT, habilitando un monitoreo remoto que resulta vital para el mantenimiento predictivo Hayman et al. (2024). En la manufactura avanzada, el mecanizado por Control Numérico Computarizado (CNC) constituye uno de los procesos más relevantes debido a su capacidad para producir componentes de alta precisión y repetibilidad. Sin embargo, a pesar de su alto nivel de automatización, la calidad final del producto depende de la estabilidad de múltiples variables físicas que interactúan de manera compleja durante el proceso de mecanizado.

Desde una perspectiva técnica, el proceso de mecanizado CNC es altamente sensible a variaciones en parámetros operacionales, los cuales presentan un comportamiento inherentemente estocástico. Esta variabilidad influye directamente en la generación de defectos, afectando tanto las propiedades superficiales como las dimensiones finales de la pieza.

Entre las variables críticas identificadas en la literatura destacan los parámetros de corte y la velocidad de avance, los cuales determinan la tasa de remoción de material. Estudios como el de Castro Ayala and Lopez Cespedes (2016) evidencian que variaciones en estos parámetros tienen una correlación directa con la rugosidad superficial y el acabado final de la pieza. Adicionalmente, enfoques recientes de monitoreo probabilístico, como los propuestos por Papananias (2025), permiten detectar anomalías en tiempo real mediante el análisis de estas variables.

Por otra parte, el uso de fluidos de corte y su presión de aplicación desempeñan un rol fundamental en la evacuación de viruta y en la disipación térmica del proceso. Investigaciones como la de Cortina et al. (2018) demuestran que una gestión inadecuada de estos fluidos puede derivar en defectos como porosidad o grietas térmicas, comprometiendo la integridad estructural del producto.

Finalmente, la temperatura del husillo y las condiciones de fricción constituyen factores críticos que afectan la precisión dimensional. Tal como señalan Cica et al. (2024), el incremento de temperatura durante el mecanizado genera dilataciones térmicas en la herramienta y la pieza, lo que se traduce en desviaciones fuera de tolerancia. La literatura confirma que la temperatura del husillo requiere un modelado termodinámico continuo, ya que es el principal causante de los errores de deformación geométrica de la pieza Feng and Yin (2014).

En este contexto, el monitoreo y modelación conjunta de estas variables adquiere un rol central. La

representación probabilística de su comportamiento, en conjunto con técnicas de análisis predictivo, permite capturar la dinámica del proceso y anticipar la aparición de defectos. De este modo, el análisis de estos parámetros físicos establece el sustento teórico para caracterizar estadísticamente el mecanizado CNC, paso fundamental para estructurar las variables de entrada de la simulación estocástica y dar inicio a la construcción del modelo.

4.4. Síntesis del Marco Teórico y Vinculación con el Problema

El marco teórico expuesto integra tres dimensiones complementarias: la gestión de activos bajo estándares internacionales como ISO 55000 y su evolución hacia el Mantenimiento Predictivo 4.0; los fundamentos probabilísticos y de ciencia de datos que permiten caracterizar la variabilidad de los procesos; y el contexto operativo de la maquinaria CNC, donde las fluctuaciones en variables críticas afectan directamente la calidad del producto.

La convergencia de estos elementos evidencia que los enfoques tradicionales de control de calidad, basados en inspección posterior y análisis descriptivo, resultan insuficientes para anticipar defectos en entornos de alta dinámica

5. Metodología

El presente capítulo describe la metodología estructurada para el diseño, desarrollo y validación del modelo probabilístico híbrido propuesto. Dada la naturaleza computacional de la investigación, se adoptó un enfoque cuantitativo y experimental basado en simulación. La solución se divide en tres grandes fases secuenciales: la modelación estocástica del entorno físico mediante un gemelo digital, la generación y preprocesamiento de la matriz de datos, y finalmente, el entrenamiento, validación y evaluación de algoritmos de clasificación supervisada. A continuación, se detallan los procedimientos algorítmicos, físicos y matemáticos aplicados en cada etapa del desarrollo.

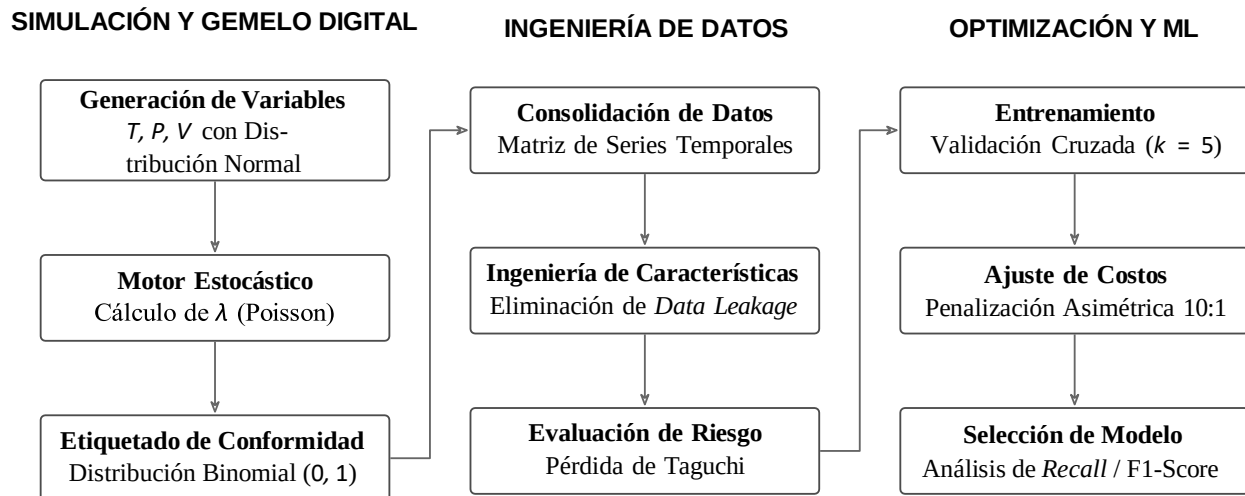


Figura 1: Diagrama metodológico completo ajustado a los márgenes de la página.

5.1. Fase 1: Modelación Estocástica del Entorno CNC

Para generar un conjunto de datos estructurado, dinámico y altamente representativo del proceso de mecanizado continuo, se diseñó un entorno de simulación computacional conceptualizado como un Gemelo Digital. Este modelo híbrido emula tanto el comportamiento físico en estado estacionario de la maquinaria (variables continuas) como la ocurrencia estocástica de fallas y defectos

(eventos discretos). La elección del entorno Simulink de MATLAB se fundamenta en su capacidad matemática para resolver sistemas de ecuaciones diferenciales en el dominio del tiempo, permitiendo sincronizar la telemetría de alta frecuencia con la lógica de control de calidad en un único flujo de trabajo.

En la Figura 2 se detalla la arquitectura de bloques desarrollada para este propósito. El flujo de datos se estructuró en tres niveles probabilísticos fundamentales que capturan el ciclo completo del mecanizado: desde la inyección de señales termomecánicas hasta el resultado final de la pieza.

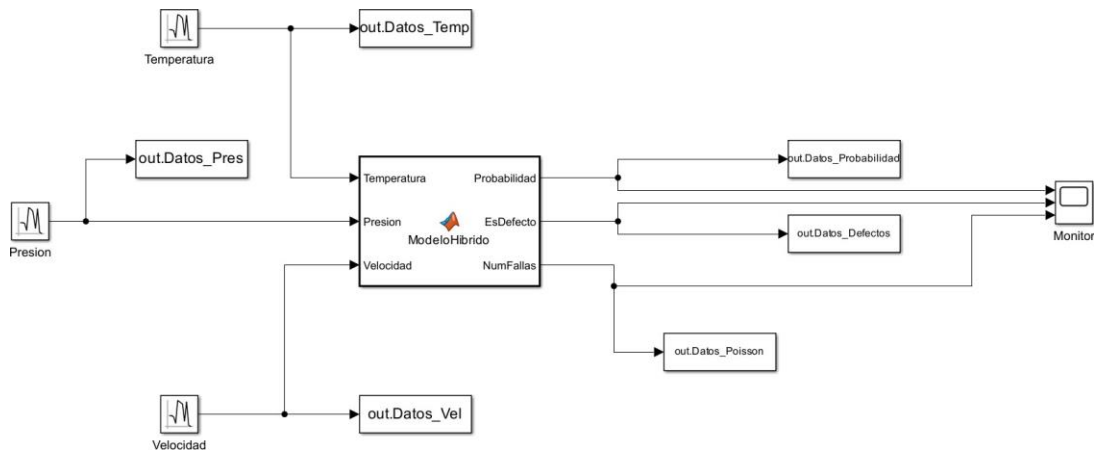


Figura 2: Arquitectura de bloques en Simulink para la simulación híbrida del gemelo digital. Se observan los bloques de entrada para temperatura, presión y velocidad, el bloque principal de 'Modelo Híbrido' que centraliza la lógica, y los bloques de salida para la probabilidad de defecto, bandera de defecto, conteo de fallas y monitorización.

5.1.1. Modelación de Variables Operacionales Continuas (Distribución Normal)

El primer nivel del modelo corresponde a la representación del comportamiento continuo de las variables físicas que gobiernan el proceso de mecanizado CNC. En particular, se consideran tres variables termomecánicas críticas: la temperatura del husillo, la presión del fluido de corte y la velocidad de operación. Estas variables reflejan directamente las condiciones bajo las cuales se

realiza la remoción de material y, por tanto, influyen de manera determinante en la calidad final de la pieza.

Desde una perspectiva ingenieril, dichas variables capturan la interacción entre fenómenos térmicos y mecánicos del proceso. La temperatura se asocia a la generación de calor por fricción, la presión del refrigerante a la capacidad de disipación térmica y evacuación de viruta, y la velocidad de operación a la estabilidad del contacto herramienta-pieza. Esta interacción da lugar a un sistema sensible a perturbaciones, donde pequeñas variaciones pueden impactar el desempeño global.

En entornos industriales, los sistemas de control automático, como PLC y variadores de frecuencia, buscan mantener estas variables en valores nominales. Sin embargo, debido a múltiples fuentes de incertidumbre, como variaciones en el material, fluctuaciones eléctricas, desgaste de la herramienta o condiciones ambientales, las variables presentan oscilaciones continuas alrededor de su punto de operación.

Este comportamiento se modela mediante una distribución normal, fundamentada en el Teorema del Límite Central, el cual establece que la suma de perturbaciones independientes tiende a una distribución gaussiana. Su función de densidad está dada por:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Donde:

- μ representa el valor promedio o central del proceso.
- σ mide la dispersión de los datos alrededor de la media.

En este modelo, μ representa el valor nominal de operación, mientras que σ cuantifica la variabilidad del proceso. Una baja dispersión indica estabilidad operativa, mientras que una mayor variabilidad puede reflejar condiciones de degradación o inestabilidad.

Desde el punto de vista computacional, la simulación incorpora ruido blanco gaussiano para introducir variabilidad controlada en las señales. Esto permite recrear tanto condiciones normales de operación como escenarios de desviación, tales como incrementos térmicos, pérdidas de presión o fluctuaciones en la velocidad.

En la simulación se define el comportamiento analógico y la línea base de los parámetros termomecánicos críticos que gobiernan el arranque de viruta: la temperatura estructural del husillo principal, la presión del fluido de corte y la velocidad de rotación. En el diagrama de bloques (Figura 2, extremo izquierdo), este estado de operación determinista se representa mediante bloques de tipo *Constant* etiquetados como T_ideal , P_ideal y V_ideal . Estos bloques inyectan los valores nominales (μ) definidos en la parametrización de ingeniería, emulando los comandos o *setpoints* estáticos que el control numérico (CNC) envía a los servomotores y bombas.

Sin embargo, en un entorno industrial real, la lectura de los sensores jamás es una línea recta. Las variables presentan fluctuaciones continuas debido a la variabilidad de la dureza del material, las vibraciones por la fricción de corte y las caídas de tensión eléctrica. Para recrear esta compleja realidad física, la implementación computacional incorpora ruido blanco gaussiano superpuesto matemáticamente a la señal ideal.

Utilizando bloques generadores de señal aleatoria (tipo *Random Number*) etiquetados como T_noise , P_noise y V_noise , el modelo inyecta perturbaciones estocásticas. Configurados bajo una Dis-

tribución Normal y parametrizados con las varianzas (σ^2) de cada sensor, estos bloques aseguran que el modelo de *Machine Learning* posterior aprenda a diferenciar entre el ruido de fondo inofensivo y las desviaciones críticas que preceden a una falla.

5.1.2. Severidad y Conteo de Anomalías (Distribución de Poisson)

El segundo nivel del modelo introduce una capa de transformación que vincula el comportamiento continuo de las variables operacionales con la generación de defectos discretos en la pieza mecanizada. Mientras que las variables del nivel anterior representan señales continuas, los defectos corresponden a eventos contables que se manifiestan en la calidad del producto, tales como irregularidades superficiales, desviaciones dimensionales o marcas de vibración.

Dado que estos eventos ocurren de forma aleatoria e independiente dentro de un intervalo de producción, su comportamiento se modela mediante una distribución de Poisson, la cual permite describir la probabilidad de ocurrencia de un número específico de defectos:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Donde:

- λ es la tasa de eventos o el promedio de defectos esperados en el intervalo.
- k es el número de eventos observados.
- e es la base del logaritmo natural.

En el modelo propuesto, este parámetro no es constante, sino que se define como una función

dinámica dependiente de las desviaciones de las variables operacionales respecto a sus valores nominales. De esta forma, cuando el proceso se mantiene estable, la tasa de defectos permanece baja; en cambio, ante desviaciones significativas, λ aumenta, reflejando un mayor nivel de deterioro en la calidad.

Este enfoque permite capturar la naturaleza no lineal del proceso, donde pequeñas variaciones pueden amplificarse y generar efectos acumulativos. En consecuencia, el modelo es capaz de representar escenarios de degradación progresiva, como el desgaste de la herramienta o la pérdida de eficiencia en el sistema de refrigeración.

En conjunto, este nivel actúa como un puente entre la variabilidad física del sistema y la manifestación observable de defectos, permitiendo traducir señales continuas en indicadores discretos de calidad.

El núcleo analítico del gemelo digital reside en un bloque de tipo *MATLAB Function* titulado `Control Calidad (Poisson)`, posicionado en el centro de la Figura 2. Este componente actúa virtualmente como un dispositivo de Controlador Lógico Programable (PLC). Su función es leer en tiempo real las señales continuas ruidosas (`T_sensor`, `P_sensor`, `V_sensor`) y transformarlas mediante lógica matemática en la aparición de anomalías discretas sobre la superficie de la pieza de trabajo.

La ocurrencia de micro-defectos (como rugosidad fuera de norma o micro-grietas) durante un ciclo de mecanizado sigue la naturaleza estadística de los eventos independientes en un intervalo de tiempo. Por ello, para cuantificar la severidad de las anomalías, se programó internamente en este bloque una función basada en la Distribución de Poisson.

La principal característica de este bloque es que su parámetro λ (la tasa promedio de ocurrencia de defectos esperados) opera como una variable dinámica e interdependiente. En cada paso de integración temporal del *solver* de Simulink, el algoritmo calcula el error cuadrático entre la lectura del sensor y el estado ideal. Si las variables físicas se alejan de su centro, λ crece, generando picos de severidad estocásticos que permiten representar fenómenos físicos no lineales, como el desgaste acelerado.

5.1.3. Etiquetado Binario del Estado del Sistema (Distribución Binomial)

El tercer nivel del modelo tiene como objetivo transformar la información generada en las etapas anteriores en una variable de salida adecuada para el entrenamiento de modelos de aprendizaje supervisado. En este contexto, la calidad del producto se representa mediante una variable binaria que clasifica cada pieza como conforme o defectuosa.

En la práctica industrial, la evaluación de calidad se basa en una decisión discreta, determinada por el cumplimiento de tolerancias específicas. Este proceso se modela mediante un ensayo de Bernoulli, caso particular de la distribución binomial cuando el número de ensayos es igual a uno:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad \text{con } n = 1$$

Donde:

- n es el número total de pruebas o unidades inspeccionadas.
- k es el número de defectos observados.
- p es la probabilidad de que ocurra un defecto en cada ensayo.

- $\binom{n}{k}$ es el coeficiente binomial, que indica las combinaciones posibles para elegir k éxitos en n intentos.

La clasificación final depende de la severidad acumulada de los defectos generados en la etapa anterior. Para ello, se establece un umbral de decisión que define el límite entre un estado aceptable y uno defectuoso. En el modelo propuesto, se define un umbral de tolerancia de $\tau = 3$ anomalías. De esta manera, si el conteo de microdefectos generado por la distribución de Poisson iguala o supera dicho valor ($k \geq 3$), la pieza es etiquetada como defectuosa; en caso contrario, se clasifica como conforme.

La selección del umbral $\tau = 3$ se fundamenta en un análisis de sensibilidad sobre la tasa de descarte del sistema. Se observó que valores menores incrementaban excesivamente la cantidad de defectos simulados, mientras que valores mayores reducían significativamente la representatividad de las fallas. Bajo esta configuración, la tasa de defectos del conjunto de datos se estabilizó en un 6.88 %, valor consistente con rangos de descarte reportados en procesos industriales de mecanizado CNC. Este procedimiento de etiquetado permite traducir el comportamiento dinámico del sistema en una variable categórica discreta, necesaria para el entrenamiento de algoritmos de clasificación supervisada. Asimismo, asegura coherencia entre la simulación probabilística desarrollada y los criterios de control de calidad utilizados en entornos industriales.

Este proceso de etiquetado permite traducir el comportamiento del sistema en una variable categórica clara, necesaria para el entrenamiento de algoritmos de clasificación. Además, asegura la coherencia entre la simulación y los criterios reales de control de calidad utilizados en la industria.

El último nivel de la arquitectura de bloques tiene como objetivo consolidar el estado físico del sistema para viabilizar su ingesta en la etapa de aprendizaje supervisado. El nivel de severidad acumulado en la etapa de Poisson (el número k de micro-defectos calculados por iteración dentro de la *MATLAB Function*) se somete internamente a un umbral de decisión estricto para generar un veredicto dicotómico (Pasa / No Pasa), modelado algorítmicamente mediante un ensayo de Bernoulli.

Si el conteo de anomalías supera la tolerancia técnica establecida, la simulación arroja un estado lógico igual a 1, etiquetando el registro instantáneo como “Pieza Defectuosa”. Por el contrario, bajo condiciones de estabilidad controlada, arroja 0 (“Pieza Conforme”). Esta variable binaria final constituye la etiqueta de clase (*Target*).

Finalmente, en el extremo derecho del modelo (Figura 2), se observan los bloques de exportación titulados `out.Datos_Probabilidad`, `out.Datos_Defectos` y `out.Datos_Poisson`. Estos componentes de tipo *To Workspace* cumplen una función crítica en la ingeniería de datos: sincronizan las series temporales generadas y las escriben directamente en la memoria de MATLAB para su posterior procesamiento y entrenamiento.

5.1.4. Parametrización y Justificación de Variables Físicas

La validez científica y la aplicabilidad industrial del modelo probabilístico dependen directamente de que los valores programados en el simulador representen de manera realista las condiciones operativas del mecanizado CNC. Un modelo basado en parámetros irreales carece de utilidad práctica, ya que conduce a predicciones que no reflejan el comportamiento del sistema físico. Por esta razón, los parámetros estadísticos base, la media nominal (μ) y la desviación estándar (σ), de cada

variable operacional se fundamentan en evidencia empírica proveniente de estudios experimentales y literatura especializada en manufactura avanzada.

Este proceso de parametrización permite asegurar que los datos sintéticos generados por el modelo mantengan coherencia con las capacidades mecánicas de la maquinaria y con las tolerancias exigidas en entornos industriales. En este sentido, no solo se busca representar valores promedio de operación, sino también capturar la variabilidad natural del proceso, elemento clave en el análisis de calidad y confiabilidad.

Los valores nominales del estado estacionario y sus respectivos niveles de dispersión se presentan en la Tabla 1. En particular, la temperatura del husillo se establece en torno a 45 [°C] con una desviación estándar de 2.5 [°C], de acuerdo con estudios experimentales sobre comportamiento térmico en sistemas de mecanizado (Arief et al., 2024). Por su parte, la presión del fluido de corte se fija en 50 [bar] con una variabilidad de 2.0 [bar], mientras que la velocidad de corte se define en 320 [m/min] con una desviación de 15 [m/min], parámetros consistentes con investigaciones sobre estabilidad del proceso y condiciones óptimas de mecanizado reportadas por Cica et al. (2024).

Variable Operacional	Símbolo	Media Nominal (μ)	Desviación Est. (σ)	Unidad
Temperatura del husillo	T	45.0	2.5	°C
Presión fluido de corte	P	50.0	2.0	bar
Velocidad de corte	V	320	15	m/min

Tabla 1: Parámetros operacionales de entrada para la simulación estocástica

Desde un punto de vista técnico, los valores de desviación estándar (σ) no representan únicamente variabilidad estadística abstracta, sino que modelan fenómenos físicos concretos presentes en el proceso. Por ejemplo, la dispersión en la velocidad de corte puede asociarse a micro-fricciones internas del sistema de transmisión, mientras que las fluctuaciones en la presión del fluido reflejan

pérdidas de carga en el circuito hidráulico. De manera similar, las variaciones en la temperatura del husillo capturan efectos acumulativos de fricción, disipación térmica y condiciones ambientales.

La incorporación de estos niveles de variabilidad permite que el modelo no solo reproduzca condiciones ideales de operación, sino también escenarios más realistas donde se manifiestan fenómenos como desgaste progresivo de la herramienta, cambios en las propiedades del material de trabajo y vibraciones mecánicas inherentes al proceso de mecanizado.

5.2. Fase 2: Generación, Procesamiento y Estructuración del Conjunto de Datos

Una vez consolidada la arquitectura estocástica en el entorno de simulación, la segunda fase metodológica se centró en la extracción, limpieza y transformación de las señales temporales continuas en datos discretos. El objetivo fundamental de esta etapa computacional fue convertir el flujo de datos dinámico y asíncrono generado por el gemelo digital en una estructura tabular estandarizada (*Dataset*), optimizada específicamente para el entrenamiento de algoritmos de aprendizaje supervisado.

El desarrollo completo de este *pipeline* de ingeniería de datos se programó íntegramente mediante la construcción de *scripts* en el entorno de MATLAB. Este proceso garantiza la trazabilidad de la información, permitiendo que la transición desde el dominio del tiempo (señales de Simulink) al dominio estadístico (matrices de datos) sea consistente y libre de errores de sincronización.

5.2.1. Extracción y Consolidación Matricial de Señales

Durante la ejecución del modelo, el *solver* de Simulink resuelve iterativamente las ecuaciones del sistema, generando series temporales de alta resolución para cada una de las variables del proceso de mecanizado. Estas señales de Temperatura, Presión, Velocidad, Severidad de fallas y Estado final, son exportadas mediante bloques `To Workspace` y almacenadas en MATLAB como arreglos multidimensionales, lo que permite conservar la evolución temporal completa del sistema.

Cada una de estas series contiene información continua indexada en el tiempo; sin embargo, los algoritmos de aprendizaje supervisado requieren estructuras tabulares donde cada observación sea independiente. Para resolver esta diferencia estructural, se implementa un proceso de preprocesamiento orientado a transformar las señales temporales en una matriz bidimensional.

Este procedimiento incluye, en primer lugar, la sincronización temporal de todas las variables, asegurando que cada valor corresponda al mismo instante de muestreo. Esta etapa es fundamental para evitar desalineaciones que puedan introducir ruido o inconsistencias en el conjunto de datos. Posteriormente, las señales son consolidadas en una estructura tipo `table`, donde cada columna representa una variable específica y cada fila corresponde a una observación discreta del sistema.

Desde una perspectiva conceptual, cada fila puede interpretarse como una “instantánea” del estado operativo de la máquina, integrando simultáneamente las condiciones del proceso y su resultado. Esta transformación permite convertir un sistema dinámico continuo en un *dataset* estructurado, preservando la información relevante y habilitando su utilización en modelos de aprendizaje automático (*Machine Learning*).

5.2.2. Ingeniería de Características y Prevención de Fuga de Información

Una de las decisiones metodológicas más críticas en la construcción del modelo predictivo es la adecuada selección de variables (feature selection) y la prevención de la fuga de información (data leakage). Este fenómeno ocurre cuando el modelo accede, durante su entrenamiento, a información que no estaría disponible en condiciones reales de operación, generando resultados artificialmente precisos pero sin validez práctica.

En la lógica del modelo de simulación, la variable de Severidad correspondiente al conteo de micro-defectos generado mediante la distribución de Poisson, actúa como un intermediario directo entre las desviaciones de las variables operacionales y el estado final de calidad. No obstante, esta variable no es observable en un entorno industrial real, donde los sistemas de monitoreo solo registran señales físicas como temperatura, presión y velocidad.

Si esta variable fuese incluida como predictor, el modelo tendería a aprender una relación directa con la salida, ignorando las variables físicas del proceso. Esto conduciría a un sobreajuste, reduciendo su capacidad de generalización y su aplicabilidad en escenarios reales.

Para evitar este problema, el proceso de preprocesamiento elimina explícitamente la variable de Severidad del dataset final. De este modo, el modelo se entrena únicamente con variables observables, manteniendo el estado binario de calidad como variable objetivo.

Adicionalmente, esta decisión refuerza la coherencia entre el modelo y su implementación práctica, ya que replica las condiciones bajo las cuales operaría un sistema real de monitoreo industrial. En consecuencia, el algoritmo se ve obligado a inferir patrones complejos a partir de las condiciones operacionales, mejorando su robustez y su capacidad de predicción frente a datos no observados.

5.2.3. Estructuración Algorítmica de la Probabilidad de Falla

Con el objetivo de dotar de realismo al proceso de generación de defectos, se implementa una función matemática que integra principios del Control Estadístico de Procesos (SPC) y de la Ingeniería de Calidad. Esta formulación permite modelar la probabilidad de falla como una variable continua dependiente de las desviaciones de las variables operacionales respecto a sus valores nominales.

En primer lugar, se establece un nivel base de probabilidad de defecto del 2 % (0.02), el cual representa la variabilidad inherente del sistema asociada a causas comunes. Incluso bajo condiciones óptimas de operación, los procesos industriales presentan un nivel residual de defectos debido a factores no controlables. La incorporación de este término permite evitar que el modelo asuma escenarios ideales irreales, mejorando su capacidad de generalización.

En segundo lugar, se modela el impacto de las desviaciones mediante la función de pérdida de Taguchi, la cual establece que el deterioro de la calidad aumenta de forma cuadrática a medida que una variable se aleja de su valor objetivo. Esta relación se expresa como:

$$L = k(x - \mu)^2 \quad (1)$$

Para extender este enfoque a múltiples variables con distintas unidades de medida, se aplica una normalización basada en la varianza, definiendo factores de ponderación de la forma $W = k/\sigma^2$. Se adopta una constante de sensibilidad empírica $k = 0,05$, lo que permite estandarizar el impacto relativo de cada variable dentro del modelo. La probabilidad de defecto para cada observación se calcula entonces mediante la siguiente expresión:

$$P_d = 0,02 + \sum_{i \in \{T, P, V\}} \frac{0,05}{\sigma_i^2} (x_i - \mu_i)^2 \quad (2)$$

donde x_i corresponde al valor observado de cada variable, mientras que μ_i y σ_i representan sus parámetros nominales y de dispersión. Esta formulación asegura que desviaciones equivalentes en términos de desviación estándar tengan un impacto comparable en la probabilidad de falla, independientemente de la escala de medición.

Desde una perspectiva metodológica, esta estructura permite generar un conjunto de datos donde la probabilidad de defecto responde de manera continua y coherente a las condiciones del proceso. De esta forma, se evita una asignación arbitraria de fallas y se logra una representación más realista del comportamiento industrial.

En síntesis, esta etapa consolida la relación entre variabilidad operativa y calidad del producto en una expresión matemática robusta, proporcionando una base consistente para el entrenamiento de modelos predictivos en la fase siguiente.

5.3. Fase 3: Entrenamiento y Evaluación de Algoritmos Predictivos

Con el conjunto de datos previamente estructurado, libre de fugas de información (*Data Leakage*) y fundamentado en criterios estadísticos, la fase final de la metodología se orienta al entrenamiento y validación de modelos de *Machine Learning*. El propósito de esta etapa es que los algoritmos sean capaces de aprender de forma autónoma los patrones subyacentes y las fronteras de decisión que vinculan las variables termomecánicas del proceso (temperatura, presión y velocidad) con la calidad final del producto (conforme o defectuoso).

Esta fase representa la transición desde la modelación del sistema hacia la construcción de un mecanismo predictivo capaz de operar bajo condiciones similares a un entorno industrial real, donde la toma de decisiones debe realizarse en tiempo oportuno y con información parcial.

5.3.1. Implementación en Entorno Classification Learner

El entrenamiento de los modelos de clasificación se implementa utilizando la herramienta *Classification Learner* del entorno MATLAB, la cual permite desarrollar, comparar y validar múltiples algoritmos de aprendizaje supervisado bajo condiciones experimentales controladas. Esta plataforma facilita la ejecución paralela de distintos modelos sobre un mismo conjunto de datos, asegurando consistencia en la evaluación y permitiendo realizar comparaciones objetivas de desempeño.

Se adopta una estrategia basada en la exploración de diversos enfoques algorítmicos, considerando que ningún modelo es universalmente óptimo. En este contexto, se selecciona un conjunto representativo de técnicas de clasificación pertenecientes a distintas familias, con el fin de capturar la complejidad del sistema:

- **Árboles de Decisión:** Segmentan los datos mediante reglas basadas en umbrales de las variables, permitiendo interpretar fácilmente cómo las condiciones operacionales influyen en la generación de defectos.
- **Máquinas de Vectores de Soporte (Quadratic SVM):** Orientadas a encontrar fronteras de separación óptimas entre clases, siendo especialmente útiles cuando las relaciones entre variables no son lineales.
- **Redes Neuronales Artificiales (MLP):** Capaces de modelar relaciones complejas entre las variables del proceso mediante la combinación de múltiples capas de procesamiento.

- **Métodos de Ensamble (Bagged Trees):** Combinan múltiples árboles de decisión para mejorar la estabilidad del modelo y reducir la variabilidad en las predicciones.

Esta diversidad de enfoques permite identificar empíricamente el modelo que mejor se adapta a la dinámica del proceso, considerando tanto su capacidad predictiva como su robustez frente a la variabilidad inherente de los sistemas industriales. En síntesis, esta etapa establece un entorno experimental estructurado que posibilita la evaluación sistemática de algoritmos de clasificación, constituyendo la base para la selección del modelo predictivo más adecuado.

5.3.2. Estrategia de Validación Cruzada

Para garantizar la confiabilidad estadística de los resultados y asegurar que el modelo seleccionado sea capaz de generalizar frente a nuevos ciclos de mecanizado, se descartó el enfoque tradicional de validación de retención simple (*hold-out*). En su lugar, se implementó un esquema de **validación cruzada de k -pliegues** (*k-fold cross-validation*), configurado con $k = 5$.

Bajo este esquema, el conjunto de datos se divide aleatoriamente en cinco subconjuntos de igual tamaño. El modelo se entrena de forma iterativa utilizando cuatro subconjuntos como datos de entrenamiento y uno como conjunto de validación. Este proceso se repite cinco veces, de modo que cada subconjunto es utilizado una vez como datos de prueba.

El desempeño final del modelo se obtiene promediando las métricas de cada iteración, lo que permite reducir la variabilidad asociada a una única partición de los datos y obtener una estimación más robusta de su rendimiento.

Este enfoque asegura que el modelo no solo aprenda patrones específicos del conjunto de entrena-

miento, sino que sea capaz de responder de manera consistente ante nuevas condiciones operacionales, lo cual es fundamental para su implementación en sistemas reales de monitoreo y control de calidad en procesos CNC.

5.3.3. Criterios de Selección: Matriz de Confusión y Costo de la No Calidad

La estrategia de entrenamiento y evaluación de los modelos no se limita a métricas globales, sino que incorpora una perspectiva económica activa mediante el Aprendizaje Sensible al Costo. En este sentido, el proceso de entrenamiento no busca únicamente maximizar la exactitud, sino que utiliza una Matriz de Costos Asimétrica para penalizar de forma diferencial los errores. Se prioriza la reducción de los falsos negativos debido a su impacto crítico en el Costo de la No Calidad (COPQ), asegurando que los algoritmos aprendan a identificar defectos incluso en condiciones de alto desbalanceo.

Esta matriz permite descomponer las predicciones en cuatro categorías fundamentales: verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN). A partir de esta estructura, es posible derivar métricas cuantitativas que permiten evaluar el rendimiento del modelo desde distintas perspectivas.

La exactitud global (*Accuracy*) mide la proporción de predicciones correctas respecto del total de observaciones, y se define como:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Si bien esta métrica entrega una visión general del desempeño, resulta insuficiente en contextos

industriales donde los tipos de error tienen impactos económicos diferenciados. Por otra parte, la precisión (*Precision*) evalúa la proporción de predicciones positivas que efectivamente corresponden a defectos reales, y se expresa como:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

Esta métrica resulta relevante para cuantificar el nivel de falsas alarmas generadas por el sistema. Sin embargo, en el contexto de la manufactura, el error más crítico corresponde a los **falsos negativos (FN)**, es decir, aquellos casos en que una pieza defectuosa es clasificada erróneamente como conforme. Este tipo de error tiene un impacto directo en el **Costo de la No Calidad (COPQ)**, ya que permite que productos fuera de especificación avancen en la cadena productiva o lleguen al cliente final.

Por esta razón, la métrica prioritaria en la selección del modelo es la sensibilidad (*Recall*), la cual mide la capacidad del algoritmo para identificar correctamente todos los defectos reales:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

Maximizar esta métrica implica minimizar la probabilidad de omitir defectos, lo cual es fundamental en sistemas de control de calidad preventiva. Adicionalmente, para complementar el análisis y equilibrar el desempeño entre precisión y sensibilidad, se considera la métrica *F1-Score*, definida como la media armónica entre *Precision* y *Recall*:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Esta métrica permite evaluar el compromiso entre detectar defectos y evitar falsas alarmas, proporcionando una visión más integral del comportamiento del modelo. En síntesis, la selección del modelo predictivo se fundamenta en un análisis multicriterio donde se prioriza la minimización de falsos negativos y la maximización de la sensibilidad.

Para materializar este enfoque preventivo, el entrenamiento de los algoritmos no se realiza bajo un criterio de error simétrico. En su lugar, se implementa una Matriz de Costos Asimétrica que penaliza los errores de clasificación de acuerdo con su impacto en la producción. Específicamente, se define una proporción de 10:1, donde el costo asignado a un Falso Negativo (omitir la detección de un defecto) es diez veces superior al de un Falso Positivo (falsa alarma).

La adopción empírica de esta magnitud se fundamenta en el principio de los costos escalonados de la calidad, frecuentemente documentado en la literatura industrial como la Regla del 10". Según establecen autores clásicos en la materia, como Gryna and Juran (1999) y Montgomery (2020), el Costo de la No Calidad (COPQ) aumenta exponencialmente a medida que un componente defectuoso avanza en la cadena de valor sin ser detectado. En el contexto de la manufactura CNC, el costo de una falsa alarma implica únicamente la detención momentánea y verificación del equipo; por el contrario, un falso negativo permite que la pieza defectuosa avance hacia el ensamblaje o llegue al usuario final, multiplicando los costos por concepto de descarte estructural, garantías y pérdida de reputación de la marca. Esta parametrización asegura que la función de optimización de los modelos priorice la Sensibilidad (Recall), alineando el comportamiento de la Inteligencia Artificial con el objetivo estratégico de proteger al cliente final.

6. Resultados

El presente capítulo expone los resultados cuantitativos obtenidos tras la ejecución de la metodología propuesta. En primer lugar, se analiza el comportamiento del gemelo digital frente a variaciones en el tiempo de simulación, evaluando su impacto en la generación estocástica de defectos. Posteriormente, se presentan y discuten las métricas de desempeño de los algoritmos de clasificación entrenados, con el objetivo de determinar el modelo más adecuado para la detección preventiva de anomalías en el proceso de mecanizado CNC.

Este análisis permite validar la eficacia de integrar modelos de pérdida de calidad de Taguchi con algoritmos de aprendizaje supervisado, proporcionando una base empírica para la toma de decisiones en entornos de manufactura avanzada.

6.1. Análisis de Sensibilidad del Horizonte de Simulación

En el contexto del modelamiento de sistemas físicos, el tiempo de simulación no constituye únicamente una variable computacional, sino un factor determinante en la representatividad estadística del proceso. Un horizonte temporal insuficiente puede impedir la manifestación de fenómenos de degradación, mientras que un horizonte adecuado permite capturar la dinámica completa del sistema, incluyendo la ocurrencia de eventos raros.

Para validar esta premisa, se diseñó un experimento iterativo en el cual el gemelo digital fue ejecutado bajo tres horizontes temporales distintos: **10, 100 y 1000 segundos** de operación continua. Esta variación controlada permite analizar cómo la exposición prolongada a la variabilidad termomecánica impacta la acumulación de desviaciones y, en consecuencia, la activación de la distribución de Poisson encargada de modelar la generación de defectos.

Para facilitar la interpretación del comportamiento estocástico en los monitores del sistema (*Scopes*) presentados en los siguientes escenarios, se definen las tres señales principales generadas por el gemelo digital:

Línea Naranja Discontinua (Severidad de Anomalías): Representa la salida discreta de la Distribución de Poisson. Muestra los “saltos” enteros que indican la cantidad de micro-defectos instantáneos generados por el estrés del mecanizado.

Línea Verde Punteada (Probabilidad de Riesgo): Refleja la fluctuación continua del riesgo dinámico, calculado mediante la suma del riesgo base por causas comunes y la función de pérdida cuadrática de Taguchi.

Línea Azul Continua (Estado de Conformidad): Corresponde a la etiqueta final binaria dictada por el ensayo de Bernoulli. Se mantiene en 0 para condiciones operativas normales y registra un salto lógico a 1 en el instante exacto en que la severidad supera la tolerancia técnica, clasificando la pieza como defectuosa.

Este análisis no solo tiene implicancias desde el punto de vista físico, sino también desde la ciencia de datos, ya que la calidad y balance del *dataset* dependen directamente del comportamiento dinámico observado durante la simulación.

6.1.1. Escenario Base: Simulación de 10 Segundos

En el primer escenario, se configuró el *solver* para un periodo de 10 segundos. Como se observa en la señal analítica del monitor del sistema (ver Figura 3), la máquina presenta únicamente

fluctuaciones leves en las variables operacionales, manteniéndose en un régimen cercano al estado nominal.

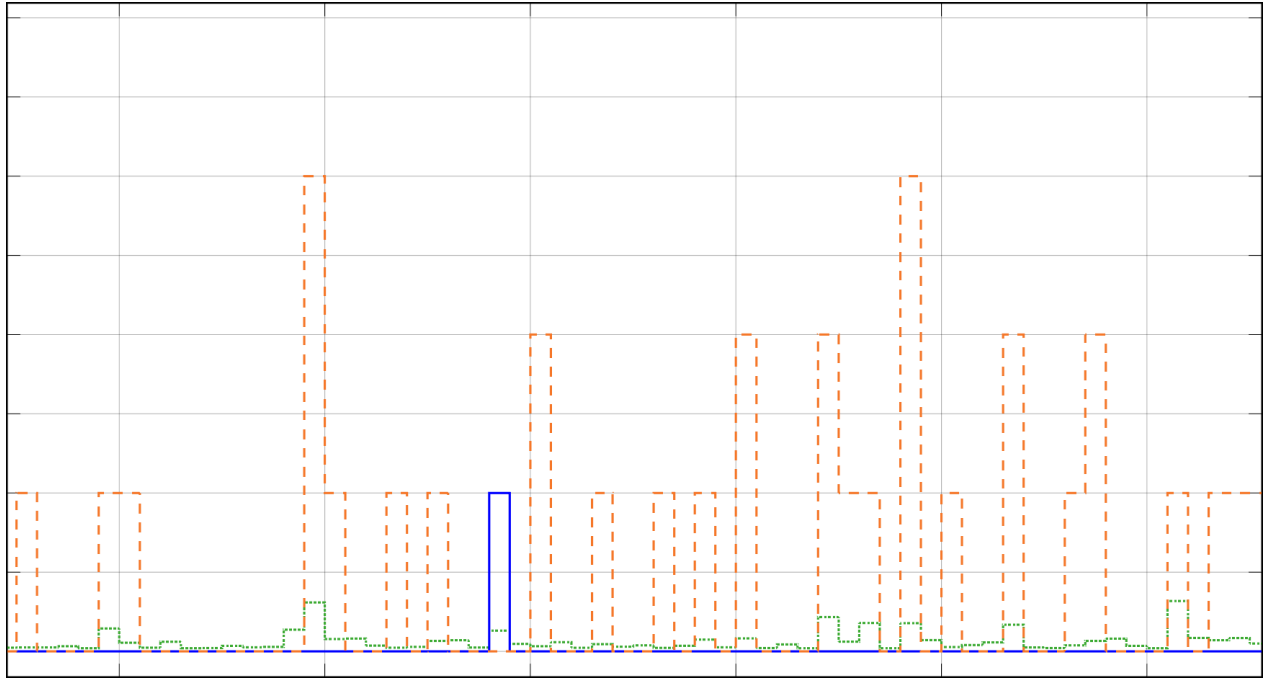


Figura 3: Comportamiento estocástico de fallas en un horizonte de 10 segundos. La línea verde punteada representa el riesgo dinámico, la naranja discontinua la severidad de anomalías y la azul continua el estado de conformidad. Se observa una baja densidad de eventos de falla, lo que limita la representatividad estadística para el entrenamiento de modelos complejos.

El conjunto de datos generado en este escenario resulta altamente desbalanceado. Debido al corto tiempo de exposición, el sistema no alcanza a acumular efectos de variabilidad significativos, por lo que la probabilidad de defecto permanece cercana a su nivel basal. Como resultado, se obtienen únicamente 1 registro defectuoso sobre un total de 101.

Este comportamiento refleja un sistema que, desde el punto de vista estadístico, no ha sido suficientemente estresado para manifestar fallas. En términos de aprendizaje automático, entrenar modelos con este tipo de datos conduce a un fenómeno de **subajuste** (*underfitting*), donde el algoritmo no logra capturar patrones relevantes y tiende a predecir mayoritariamente la clase dominante (piezas

conformes).

6.1.2. Escenario Intermedio: Simulación de 100 Segundos

Al incrementar el horizonte de simulación a 100 segundos, el modelo computacional comienza a capturar de manera más evidente la dinámica de degradación del sistema. En este intervalo, las variables termomecánicas presentan una mayor dispersión respecto a sus valores nominales, lo que se traduce en un aumento progresivo de la probabilidad de falla. La Figura 4 evidencia una mayor densidad en los picos de severidad (línea naranja), así como una tendencia creciente en la señal de riesgo asociada al proceso.

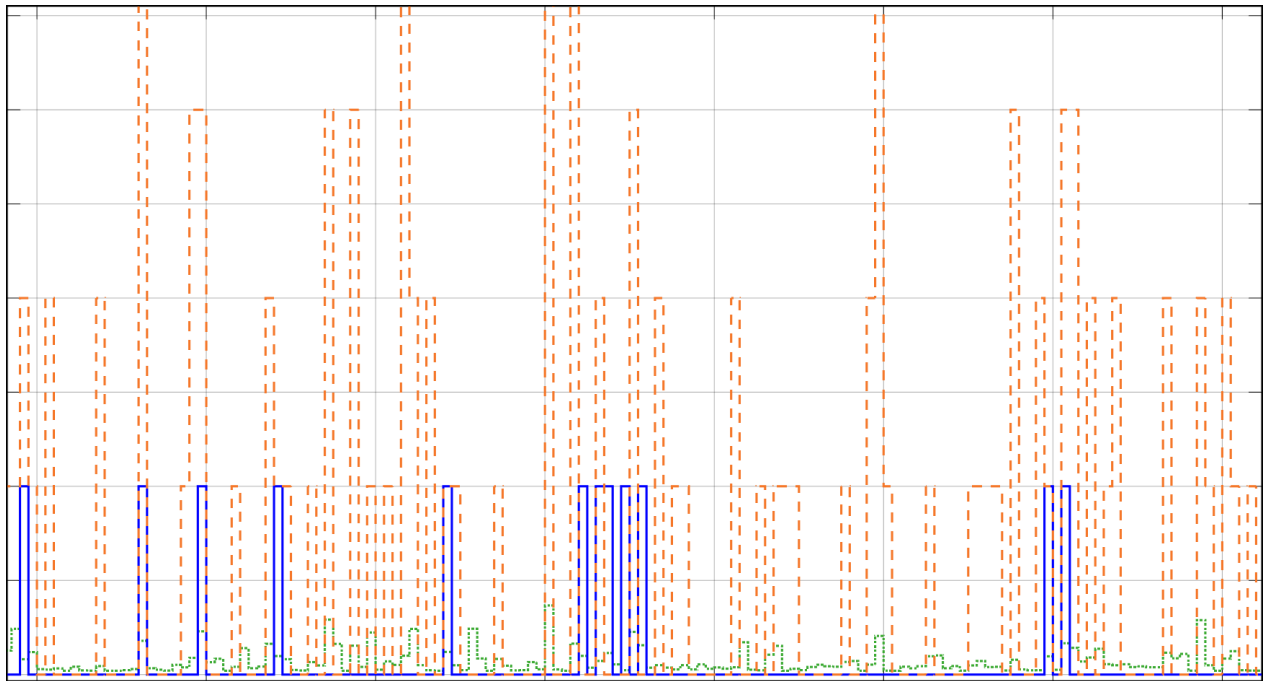


Figura 4: Evolución temporal de anomalías en un horizonte de 100 segundos. La línea verde punteada representa el riesgo dinámico; la línea naranja discontinua indica la severidad de anomalías (λ de Poisson); y la línea azul continua en forma de escalón señala el estado binario de conformidad (Bernoulli). En este intervalo se aprecia un incremento en la frecuencia de defectos respecto al escenario base.

Como resultado de este escenario, el conjunto de datos generado alcanzó un total de 1001 registros,

de los cuales 61 fueron clasificados como piezas defectuosas. Este incremento en la cantidad de eventos discretos se debe a que el mayor tiempo de exposición permite la acumulación de desviaciones en las variables operacionales, activando con mayor frecuencia la distribución de Poisson que modela la ocurrencia de anomalías.

Desde el punto de vista estadístico, este *dataset* presenta una mejor representación de ambas clases en comparación con el escenario base de 10 segundos, reduciendo parcialmente el problema de desbalanceo. No obstante, la proporción de defectos aún puede considerarse limitada para el entrenamiento de modelos más complejos, especialmente aquellos con alta capacidad de aprendizaje como las redes neuronales, los cuales requieren una mayor cantidad de ejemplos para capturar patrones no lineales de manera robusta.

En consecuencia, este escenario constituye un punto intermedio que evidencia la transición entre un sistema subrepresentado y uno estadísticamente significativo, destacando la importancia del horizonte temporal en la generación de datos sintéticos para aplicaciones de aprendizaje supervisado.

6.1.3. Escenario Óptimo: Simulación de 1000 Segundos

Finalmente, la simulación se extiende a 1000 segundos, lo que permite capturar de manera completa el comportamiento estocástico del sistema bajo condiciones prolongadas de operación. Como se observa en la Figura 5, este horizonte temporal permite visualizar una dinámica mucho más rica en términos de la interrelación entre las variables físicas y la generación de fallas.

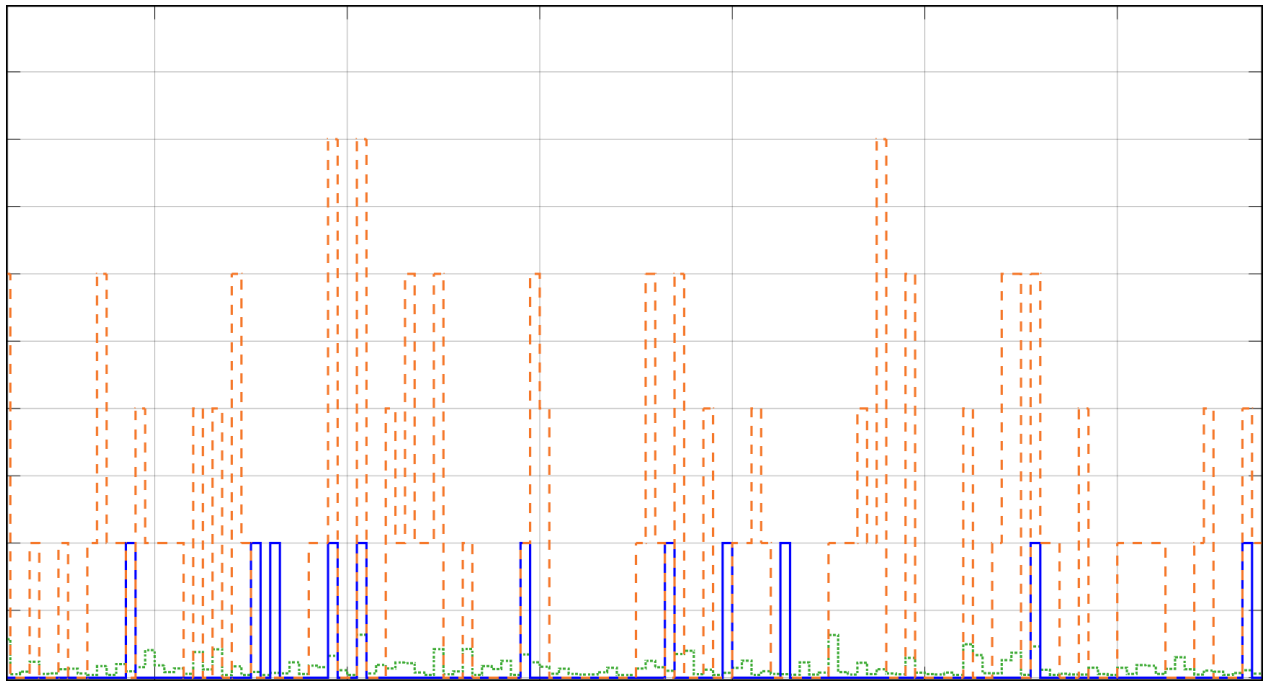


Figura 5: Consolidación probabilística y ocurrencia de defectos en un régimen de 1000 segundos (vista ampliada). La línea verde punteada representa la probabilidad de riesgo dinámico; la línea naranja discontinua indica la severidad de anomalías (λ de Poisson); y la línea azul continua en forma de escalón señala el estado binario de conformidad, donde el valor 1 representa la ocurrencia de un defecto (Bernoulli).

En este escenario, el conjunto de datos alcanza un total de 10001 registros, dentro de los cuales se identifican 688 eventos de no conformidad. La mayor duración de la simulación permite que las variables operacionales experimenten ciclos completos de variabilidad, incluyendo desviaciones acumulativas y eventos transitorios que incrementan la probabilidad de falla.

Desde el punto de vista estadístico, este escenario logra capturar adecuadamente la naturaleza de los **eventos raros**, característica fundamental en procesos industriales donde los defectos no ocurren de manera frecuente, pero tienen un alto impacto. Además, el *dataset* resultante presenta un mejor balance entre clases, lo que favorece el entrenamiento de modelos predictivos más robustos.

En términos metodológicos, este horizonte de simulación proporciona una base de datos suficiente-

mente amplia y representativa, permitiendo que los algoritmos de clasificación aprendan patrones significativos y generalicen de manera efectiva frente a nuevos datos. Por esta razón, se establece que el horizonte de 1000 segundos constituye la configuración óptima para el desarrollo del modelo en esta investigación.

6.2. Evaluación de Modelos bajo Aprendizaje Sensible al Costo

Una vez estructurado el conjunto de datos bajo el horizonte temporal óptimo de 1000 segundos, se procede al entrenamiento de los modelos predictivos. Debido al desbalanceo intrínseco de las clases, los algoritmos tradicionales tienden a sobreajustarse hacia la clase mayoritaria, alcanzando altos niveles de exactitud global, pero con una capacidad limitada para detectar defectos reales.

Con el fin de mitigar esta limitación y alinear el desempeño de los modelos con los objetivos de la gestión de calidad, se implementa una estrategia de **aprendizaje sensible al costo** (*Cost-Sensitive Learning*). En el contexto de la manufactura, el impacto asociado a los **falsos negativos** resulta significativamente mayor que el de los falsos positivos, debido a las consecuencias operativas y económicas derivadas de la no detección de defectos. En consecuencia, se define una matriz de costos asimétrica, estableciendo una penalización en proporción 10:1 para los errores asociados a defectos no detectados.

Esta configuración permite reorientar el proceso de entrenamiento hacia la maximización de la capacidad de detección, priorizando la sensibilidad del modelo por sobre la exactitud global. Los resultados obtenidos, validados mediante un esquema de validación cruzada de 5 pliegues, se presentan en la Tabla 2, considerando métricas de desempeño relevantes como sensibilidad (*Recall*), precisión y *F1-Score*, las cuales permiten evaluar el equilibrio entre detección efectiva y control de

falsas alarmas (ver Figura 6).

Modelo	Accuracy	Recall	Precision	F1-Score
Árbol de Decisión	78.3 %	50.4 %	15.9 %	24.2 %
SVM	54.1 %	31.8 %	5.0 %	8.6 %
Red Neuronal	75.7 %	61.2 %	16.3 %	25.7 %
Random Forest	85.8 %	18.5 %	12.8 %	15.1 %

Tabla 2: Desempeño de modelos bajo aprendizaje sensible al costo (Relación 10:1).

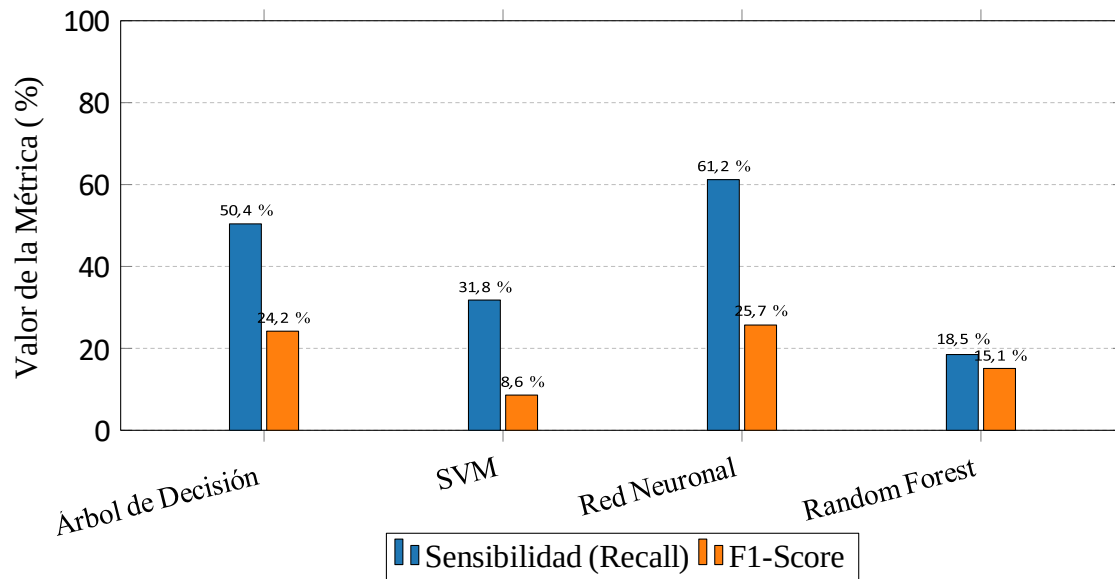


Figura 6: Comparación de métricas de desempeño bajo aprendizaje sensible al costo (Relación 10:1). Se observa que la Red Neuronal logra el mayor Recall (61.2 %), priorizando la detección de defectos críticos.

El análisis general de los resultados bajo la directriz de costos asimétricos evidencia el perfil de rendimiento definitivo de los modelos. Se observa un nivel de exactitud global moderado, el cual se asume como un compromiso (trade-off) necesario para lograr una sensibilidad operativamente viable. Esta configuración estructural refleja la prioridad gerencial de maximizar la detección preventiva de defectos por sobre la reducción de falsas alarmas.

Asimismo, se identifica que los modelos con mayor capacidad para adaptarse a funciones de pérdi-

da complejas, como las redes neuronales, logran responder de manera más efectiva a este esquema de penalización. En conjunto, estos resultados demuestran que la incorporación de criterios económicos en el proceso de entrenamiento permite alinear las herramientas de aprendizaje automático con los objetivos estratégicos de la gestión de calidad en manufactura avanzada.

6.2.1. Desempeño de la Red Neuronal Artificial

Bajo la configuración de penalización asimétrica, la Red Neuronal Artificial presenta el mejor desempeño relativo entre los modelos evaluados. Su arquitectura no lineal, basada en múltiples capas de procesamiento, permite capturar relaciones complejas entre las variables operacionales del proceso, particularmente aquellas asociadas a interacciones no evidentes entre temperatura, presión y velocidad.

En términos de desempeño, el modelo alcanza una **sensibilidad (Recall) del 61.2 %**, lo que indica una capacidad significativa para detectar defectos reales generados en el proceso de mecanizado. Este resultado adquiere especial relevancia en el contexto industrial, donde la identificación temprana de anomalías permite reducir el impacto del costo de no calidad (COPQ) y prevenir la propagación de fallas en etapas posteriores de producción.

Adicionalmente, la red neuronal mantiene una **exactitud global (Accuracy) del 75.7 %**, lo que refleja un comportamiento estable frente al conjunto total de datos, incluso bajo condiciones de desbalanceo. Esta combinación de alta capacidad de detección y desempeño general adecuado evidencia una correcta adaptación del modelo a la estructura del problema.

Finalmente, el modelo alcanza el **mayor valor de F1-Score (25.7 %)** entre las alternativas evalua-

das, lo que confirma su capacidad para equilibrar la detección de defectos con el control de falsas alarmas. En conjunto, estos resultados posicionan a la Red Neuronal Artificial como una solución robusta y adecuada para su implementación en sistemas de monitoreo predictivo en entornos de manufactura avanzada.

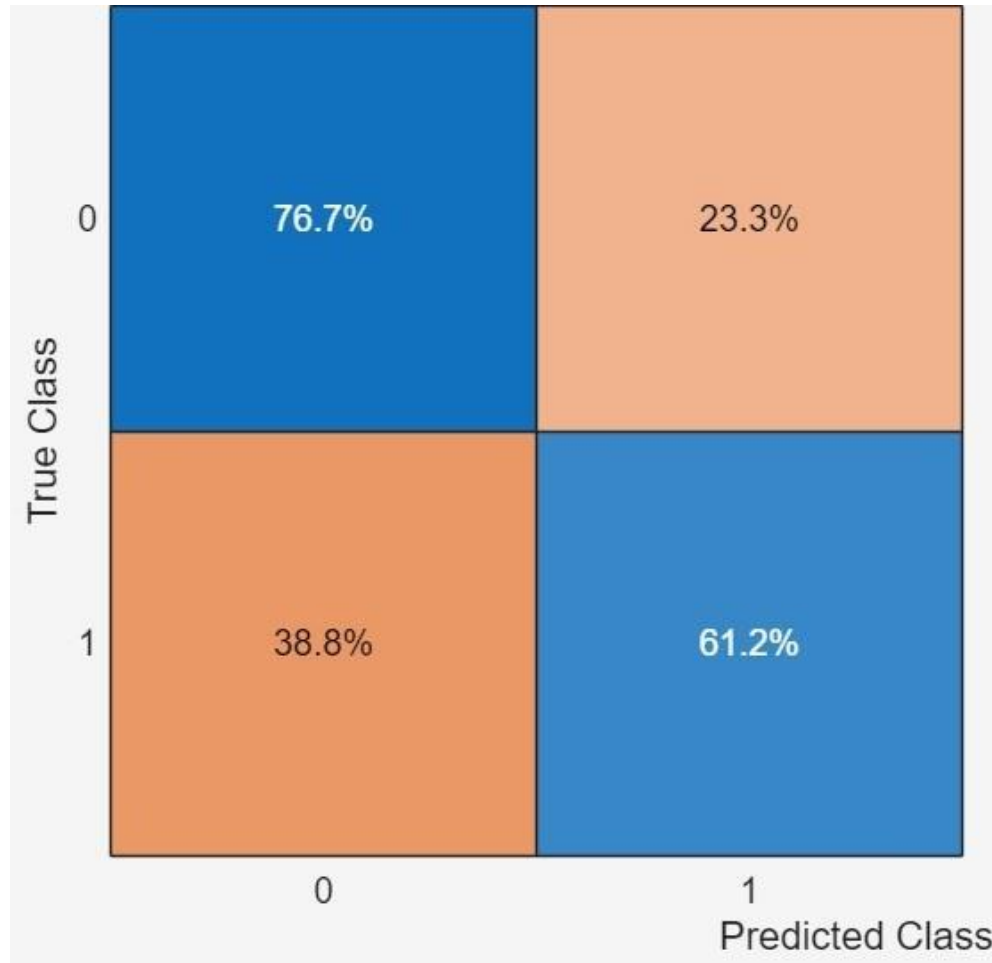


Figura 7: Matriz de Confusión de la Red Neuronal bajo aprendizaje sensible al costo.

6.2.2. Evaluación de Modelos Secundarios

Al evaluar el desempeño de los modelos restantes bajo la matriz de costos asimétrica (10:1), se observan comportamientos diferenciados que reflejan limitaciones estructurales en su capacidad de adaptación a esquemas de penalización orientados a la detección de defectos.

En el caso del modelo **Quadratic SVM**, se evidencia un desempeño global deficiente bajo este esquema. Impulsado por la penalización asociada a los falsos negativos, el algoritmo ajusta su frontera de decisión con el objetivo de maximizar la detección de la clase minoritaria. Sin embargo, este ajuste genera un desplazamiento excesivo del hiperplano de separación, afectando negativamente su capacidad de discriminación. Como resultado, la exactitud global se sitúa en un crítico (**54.1 %**) y la precisión alcanza un valor de (**5.0 %**), lo que indica una alta proporción de falsas alarmas. Desde una perspectiva operativa, este comportamiento compromete la utilidad del modelo, ya que induce intervenciones innecesarias que afectan la continuidad del proceso productivo.

Por otra parte, el método de ensamble basado en **Bagged Trees (Random Forest)** presenta un comportamiento opuesto. Debido a su estructura basada en la agregación de múltiples árboles entrenados sobre subconjuntos aleatorios, el modelo tiende a estabilizar sus predicciones mediante el promedio, reduciendo la varianza. No obstante, esta misma característica limita su capacidad de respuesta frente a la penalización asimétrica. En efecto, el modelo mantiene una alta exactitud global (**85.8 %**), pero presenta una sensibilidad reducida (**18.5 %**), lo que indica una baja capacidad para detectar defectos reales.

Este comportamiento sugiere que el ensamble prioriza la consistencia estadística de las predicciones por sobre la adaptación a la estructura de costos del problema, lo que resulta en una estrategia conservadora que permite la persistencia de errores críticos. En contextos industriales donde la detección temprana de fallas es prioritaria, esta limitación reduce su aplicabilidad práctica. En conjunto, estos resultados evidencian que la respuesta de los modelos frente a esquemas de aprendizaje sensible al costo no es homogénea, y que la elección del algoritmo debe considerar no solo su desempeño global, sino también su capacidad de adaptación a criterios operacionales específicos.

COS.

6.3. Análisis

A partir de los resultados obtenidos en la evaluación comparativa de modelos bajo un esquema de aprendizaje sensible al costo, es posible establecer una jerarquía clara de desempeño en función de los objetivos del sistema propuesto. En particular, la evaluación bajo una matriz de costos asimétrica demuestra de manera empírica que el comportamiento de los algoritmos y la selección del modelo óptimo no dependen exclusivamente de métricas tradicionales, sino de su alineación estricta con los criterios operacionales del entorno industrial.

En este contexto, la **Red Neuronal Artificial** se posiciona como la alternativa más adecuada para la implementación del sistema predictivo. Su desempeño refleja una adecuada capacidad para modelar relaciones no lineales entre las variables termomecánicas del proceso, lo que le permite identificar patrones complejos asociados a la generación de defectos. Asimismo, su mayor nivel de sensibilidad en comparación con los demás modelos evidencia una mejor adaptación al objetivo de detección preventiva.

Desde una perspectiva operativa, la capacidad del modelo para identificar una proporción significativa de fallas incipientes implica una reducción directa del Costo de la No Calidad (COPQ), al evitar la propagación de defectos en etapas posteriores del proceso productivo. Al mismo tiempo, su nivel de precisión y exactitud global se mantiene dentro de rangos aceptables, lo que permite su implementación sin comprometer de manera significativa la continuidad operativa del sistema.

Es importante destacar que los valores absolutos de F1-Score obtenidos (máximo 25.7 %) y la sen-

sibilidad del 61.2 % pueden parecer reducidos en un contexto informático tradicional. Sin embargo, en un entorno industrial altamente desbalanceado (donde los defectos representan menos del 7 % de la muestra total), estos valores son esperables y reflejan el trade-off matemático forzado por la matriz de costos 10:1. El modelo sacrifica intencionalmente la Precisión (aumentando la tasa de falsas alarmas) para lograr un Recall viable que proteja de manera efectiva la cadena de valor.

Por otra parte, el análisis de los modelos alternativos confirma que un alto nivel de exactitud global no garantiza un desempeño adecuado en contextos industriales. En particular, modelos con baja sensibilidad tienden a subestimar la ocurrencia de defectos, mientras que aquellos que sobrerreaccionan a la penalización generan un volumen excesivo de falsas alarmas. Estos resultados refuerzan la importancia de incorporar criterios económicos y operacionales en la fase de entrenamiento.

En consecuencia, la selección de la Red Neuronal Artificial como motor analítico del gemelo digital propuesto se encuentra plenamente justificada tanto desde el punto de vista estadístico como desde su aplicabilidad industrial. Este resultado permite validar la hipótesis central de la investigación, la cual establece que la efectividad de los sistemas predictivos en manufactura avanzada depende no solo de la disponibilidad de datos o de la complejidad del modelo, sino de su adecuada calibración en función de los costos asociados al proceso.

7. Conclusiones

El presente capítulo consolida los principales hallazgos y reflexiones derivadas del desarrollo de la investigación. Considerando la variabilidad estocástica inherente a los procesos de manufactura avanzada y su impacto sobre la calidad del producto y los costos operacionales, el trabajo desarrollado propone una transición metodológica desde enfoques reactivos tradicionales hacia un esquema de control predictivo basado en datos y herramientas de modelación computacional. En este contexto, se presenta una discusión de los resultados obtenidos, el impacto conceptual del modelo propuesto y las principales proyecciones para futuras investigaciones.

7.1. Discusión de resultados

En relación con los objetivos planteados al inicio de la investigación, los resultados obtenidos permiten establecer que estos fueron cumplidos satisfactoriamente, evidenciando una coherencia metodológica entre la formulación teórica, el desarrollo computacional y la evaluación experimental del modelo.

El **primer objetivo específico** se aborda mediante la caracterización probabilística de las variables termomecánicas críticas del proceso de mecanizado CNC, específicamente temperatura, presión y velocidad. La modelación de estas variables a través de distribuciones normales permite representar de manera consistente la variabilidad operacional inherente al sistema, capturando las fluctuaciones naturales presentes en un entorno industrial.

El **segundo objetivo** se cumple mediante el desarrollo de un entorno de simulación estocástica inspirado en la lógica de los gemelos digitales. Este entorno permite generar un conjunto de datos sintéticos compuesto por 10.001 registros operacionales, modelando de manera coherente la

ocurrencia discreta de defectos mediante una distribución de Poisson. La arquitectura desarrollada demuestra que es posible representar computacionalmente la degradación progresiva del proceso y la acumulación probabilística de anomalías sin depender exclusivamente de datos históricos reales.

El **tercer objetivo** se materializa mediante el entrenamiento y comparación de distintos algoritmos de clasificación supervisada, incluyendo Árboles de Decisión, Máquinas de Vectores de Soporte, Métodos de Ensamble y Redes Neuronales Artificiales. Los resultados evidencian que los modelos de aprendizaje supervisado son capaces de identificar patrones relevantes en los datos generados, incluso bajo condiciones de ruido y desbalance de clases.

Por otra parte, el **cuarto objetivo** se satisface mediante la implementación de una estrategia de aprendizaje sensible al costo. La incorporación de una matriz de costos asimétrica (10:1) permite reorientar el proceso de aprendizaje hacia la detección prioritaria de la clase minoritaria, asociada a las piezas defectuosas. Este enfoque resulta particularmente relevante en contextos industriales, donde la no detección de un defecto implica consecuencias operacionales y económicas considerablemente más críticas que una falsa alarma preventiva.

Desde el punto de vista cuantitativo, la Red Neuronal Artificial presenta el mejor equilibrio entre sensibilidad y estabilidad operativa. Bajo el esquema de penalización asimétrica, el modelo alcanza un **Recall del 61.2 %** para la clase defectuosa y un **F1-Score de 25.7 %**, manteniendo una **Exactitud Global de 75.7 %**. Si bien el valor del F1-Score permanece moderado debido al fuerte desbalance de clases presente en el conjunto de datos, el incremento de la sensibilidad evidencia una mejora significativa en la capacidad de detección preventiva de defectos críticos.

En este sentido, los resultados obtenidos evidencian que es posible incrementar la capacidad de

detección preventiva mediante una reducción controlada de la precisión global del modelo. Esta relación entre sensibilidad y exactitud refleja el compromiso inherente a los sistemas predictivos aplicados a manufactura avanzada, donde la prioridad operacional se centra en minimizar el riesgo asociado a la liberación de piezas defectuosas.

En consecuencia, los resultados obtenidos permiten aceptar la hipótesis de trabajo planteada, evidenciando que la integración de simulación estocástica y aprendizaje supervisado, bajo un enfoque sensible al costo, mejora la capacidad de detección preventiva de defectos en comparación con enfoques estadísticos descriptivos tradicionales.

7.2. Impacto estratégico y operacional

El aporte de la presente investigación trasciende el ámbito estrictamente técnico, situándose en la conceptualización de un paradigma predictivo alineado con los principios de la Industria 4.0 y la Calidad 4.0. A diferencia de metodologías convencionales que operan bajo una lógica reactiva, el modelo desarrollado evidencia el potencial teórico-computacional de anticipar desviaciones operacionales durante la ejecución del proceso productivo.

En este contexto, la integración de simulación estocástica y aprendizaje automático permite trasladar el enfoque de control desde la inspección final hacia el monitoreo continuo de variables operacionales críticas. Esta transición representa un cambio relevante desde el punto de vista de la gestión de calidad, ya que posibilita identificar condiciones anómalas antes de que estas se traduzcan en productos no conformes.

Asimismo, la incorporación de una estrategia de aprendizaje sensible al costo introduce una dimen-

sión económica directamente asociada a la toma de decisiones del modelo predictivo. Al penalizar de manera diferenciada los falsos negativos, el sistema prioriza la detección de defectos críticos por sobre la maximización exclusiva de la exactitud global. Este comportamiento resulta coherente con los principios de la gestión industrial, donde la detección tardía de una falla genera costos considerablemente superiores asociados a reprocesos, desperdicio de material y pérdida de confiabilidad operacional.

Desde una perspectiva conceptual y bajo un entorno de simulación, los resultados obtenidos sugieren una capacidad potencial para contribuir a la reducción del Costo de la No Calidad (COPQ). Si bien el modelo no ha sido implementado en una planta industrial real, la simulación evidencia computacionalmente que anticipar la falla resulta operacionalmente más conveniente que permitir el avance de piezas defectuosas dentro de la cadena productiva.

De este modo, el modelo desarrollado no solo constituye una aproximación computacional para la detección de defectos, sino también una base conceptual para futuras arquitecturas predictivas aplicadas a sistemas de manufactura inteligente y toma de decisiones autónoma basada en datos.

7.3. Limitaciones y trabajos futuros

A pesar de la consistencia de los resultados obtenidos, la presente investigación se desarrolla en un entorno de simulación computacional, lo que implica limitaciones respecto a su validación empírica. Si bien el gemelo digital permite representar de manera controlada la variabilidad del proceso, la ausencia de sincronización IoT en tiempo real y retroalimentación física restringe la verificación directa del modelo bajo condiciones operacionales reales.

En primer lugar, el sistema desarrollado se fundamenta en variables termomecánicas de respuesta relativamente lenta, tales como temperatura, presión y velocidad. Aunque estas variables resultan representativas del comportamiento general del proceso, no capturan fenómenos dinámicos de alta frecuencia asociados a eventos críticos, como vibraciones mecánicas, emisión acústica del husillo o microvariaciones instantáneas de fricción. La incorporación de este tipo de señales en futuras investigaciones podría incrementar significativamente la sensibilidad y capacidad predictiva de los algoritmos.

Asimismo, el modelo propuesto posee un carácter fundamentalmente predictivo, orientado a la detección de anomalías. Una extensión natural de esta investigación corresponde al desarrollo de arquitecturas prescriptivas capaces de integrar la salida del modelo con sistemas de control automático. Bajo este enfoque, el sistema no solo identificaría condiciones de riesgo, sino que además podría ejecutar acciones correctivas autónomas sobre los parámetros de operación de la máquina.

Finalmente, se identifica como línea de desarrollo futura la validación experimental del modelo en entornos industriales reales mediante arquitecturas basadas en Internet de las Cosas (IoT) y sistemas de adquisición de datos en tiempo real. Esta integración permitiría contrastar las distribuciones probabilísticas utilizadas en la simulación con el comportamiento efectivo de maquinaria CNC operando bajo condiciones reales de producción.

En conjunto, estas líneas de trabajo representan oportunidades concretas para ampliar el alcance y aplicabilidad del modelo desarrollado, constituyendo una base metodológica para el desarrollo de futuras arquitecturas predictivas orientadas a entornos de manufactura inteligente.

Referencias

- Arief, T. M., Lin, W.-Z., Royandi, M. A., and Hung, J.-P. (2024). Development of a digital model for predicting the variation in bearing preload and dynamic characteristics of a milling spindle under thermal effects. *Lubricants*, 12(6):185.
- Biard, G. and Nour, G. A. (2021). Industry 4.0 contribution to asset management in the electrical industry. *Sustainability*, 13(18):10369.
- Bratić, D., Miljković, P., Jurečić, D., and Grabarić, T. (2025). Ai-driven random forest model and the six sigma approach for enhancing offset printing process and product quality. *Applied Sciences*, 15(5):2266.
- Campuzano, W. P. R., Salazar, J. A. P., and Ramos, M. M. M. (2022). Sistema automático de adquisición y análisis de datos con machine learning para el control de calidad en una máquina de inyección de plástico. *Polo del Conocimiento: Revista científico-profesional*, 7(2):4.
- Castro Ayala, H. D. and Lopez Cespedes, L. C. (2016). Optimización de los parámetros de corte en la rugosidad en un acero aisi sae 1045.
- Cica, D., Sredanovic, B., Tesic, S., and Kramar, D. (2024). Predictive modeling of turning operations under different cooling/lubricating conditions for sustainable manufacturing with machine learning techniques. *Applied Computing and Informatics*, 20(1-2):162–180.
- Cortina, M., Arrate, J. I. A., Ruiz, J. E., Arrien, E. U., and Mentxaka, A. L. (2018). Análisis de la influencia del uso de fluido de corte en procesos híbridos de mecanizado y aporte por láser. *Revista iberoamericana de ingeniería mecánica*, 22(2):33–39.

- Feng, Z. and Yin, G. (2014). Variable-weight combination prediction of thermal error modeling on cnc machine tools. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 12(9):6797–6804.
- Garza Ríos, R. C., González Sánchez, C. N., Rodríguez González, E. L., and Hernández Asco, C. M. (2016). Aplicación de la metodología dmaic de seis sigma con simulación discreta y técnicas multicriterio. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 22:19–35.
- Gryna, F. M. and Juran, J. M. (1999). *Juran's Quality Handbook*. McGraw-Hill.
- Guo, M., Wei, S., Han, C., Xia, W., Luo, C., and Lin, Z. (2024). Prediction of surface roughness using deep learning and data augmentation. *Journal of Intelligent Manufacturing and Special Equipment*, 5(1):221–241.
- Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.
- Hayman, L. H. S., Tamayo, Y. M., and Ronquillo, D. F. T. (2024). Incidencia de la tecnología cnc en la industria 4.0. *Revista Ingenio global*, 3(2):37–54.
- Hernández-Montero, F. E., Anías-Calderón, C., and Ruiz-Barrios, M. L. (2023). El camino hacia la implementación del mantenimiento predictivo 4.0 en cuba. *Ingeniería Industrial*, 44(3):17–28.
- Lin, C.-Y., Wu, S.-C., Hu, C.-H., and Jian, B.-L. (2023). Spindle thermal error prediction based on back propagation neural network and hybrid taguchi genetic algorithm for a computer numerical control machine tool. *Sens. Mater*, 35(12):4397–4415.

- Llanes-Font, M. and Lorenzo-Llanes, E. (2021). La cuarta revolución industrial y una nueva aliada: calidad 4.0. *Ciencias Holguín*, 27(2):67–78.
- Lloor, J. J. M., Sánchez, M. A. J., Cuadrado, P. A. V., and Vivanco, D. C. E. (2025). Mantenimiento predictivo con inteligencia artificial en equipos electromecánicos para minimizar riesgos ambientales. *Revista Científica Arbitrada Multidisciplinaria PENTACIENCIAS*, 7(5):482–493.
- Montgomery, D. C. (2020). *Introduction to statistical quality control*. John wiley & sons.
- Papananias, M. (2025). Bayesian monitoring of machining processes using non-intrusive sensing and on-machine comparator measurement. *The International Journal of Advanced Manufacturing Technology*, 137(3):1929–1942.
- Pérez, L. P. T., Guamán, J. S. B., and Sampedro, J. E. V. (2022). Simulación de un modelo matemático con matlab (simulink), para dos tanques industriales conectado en paralelo. *Polo del Conocimiento*, 7(6):2652–2667.
- Rabanal, L. J. C., Gordillo, A. S. H., and Rengifo, M. A. L. (2024). El impacto de la aplicación de la tecnología de gemelos digitales en la cadena de suministro 4.0: Una revisión sistemática. *Gestión De Operaciones Industriales*, 3(1):80–92.
- Salazar-Zuluaga, A., Zapata-Madrigal, G. D., and García Sierra, R. (2021). Electrofun: una aplicación basada en gamificación para apoyar la implementación del sistema de gestión de activos iso 55000 en codensa. *Ingeniería y Desarrollo*, 39(1):138–155.
- Shuriya, B. (2026). Ai-powered predictive maintenance via sensor fusion and machine learning in downtime reduction and equipment efficiency in industry 4.0 manufacturing plants.

- Toro Lazo, A. (2024). El enfoque predictivo de la inteligencia artificial como herramienta para el mantenimiento de maquinaria industrial: una mirada desde la industria 4.0. *Entre ciencia e ingeniería*, 18(35):7–8.
- Walker, C., Luo, X., Kundu, P., Chang, W., Xie, W., Ball, P., and Badakhshan, E. (2022). Digital twin technology for cnc machining: a review. In *2022 8th International Conference on Nanomanufacturing & 4th AET Symposium on ACSM and Digital Manufacturing (Nanoman-AETS)*, pages 1–6. IEEE.
- Xiao, G., Tu, H., Xu, Y., Shao, J., and Xiang, D. (2024). Surface roughness prediction based on cnn-bitcn-attention in end milling. *International Journal of Advanced Computer Science & Applications*, 15(12).
- Zhang, Z., Lv, X., Qi, B., Qi, Y., Zhang, M., and Tao, Z. (2024). Surface roughness prediction and roughness reliability evaluation of cnc milling based on surface topography simulation. *Eksploatacja i Niezawodność*, 26(2).