

**DESARROLLO DE PROTOTIPO DE SISTEMA INTELIGENTE PARA REGISTRO
E IDENTIFICACIÓN DE MARIPOSAS EN EL PARQUE METROPOLITANO
MARIA LUCIA (VILLAVICENCIO) BASADO EN HMM Y RNA**

AUTOR

OSCAR JOBANY GOMEZ OCHOA

**UNIVERSIDAD DE CUNDINAMARCA
FACULTAD DE INGENIERÍA
PROGRAMA INGENIERÍA DE SISTEMAS
FACATATIVÁ, MAYO DE 2018**

**DESARROLLO DE PROTOTIPO DE SISTEMA INTELIGENTE PARA REGISTRO
E IDENTIFICACIÓN DE MARIPOSAS EN EL PARQUE METROPOLITANO
MARIA LUCIA (VILLAVICENCIO) BASADO EN HMM Y RNA**

AUTOR

OSCAR JOBANY GOMEZ OCHOA

**TRABAJO DE GRADO PRESENTADO COMO REQUISITO PARA OPTAR AL
TITULO DE INGENIERO DE SISTEMAS**

DIRECTOR: ING. JAIME EDUARDO ANDRADE RAMÍREZ

CODIRECTORA: LIC. LUZ DARY PEDRAZA HERNÁNDEZ

**GRUPO DE INVESTIGACIÓN DE SISTEMAS Y TECNOLOGÍA DE FACATATIVÁ
(GISTFA)**

CENTRO DE INVESTIGACION Y TECNOLOGIA (CIT) DE FACATATIVÁ

SEMILLERO DE ROBOTOTICA FACATATIVA (SERFAC)

CENTRO DE INVESTIGACIONES AMBIENTALES (CIAM) DE VILLAVICENCIO

**UNIVERSIDAD DE CUNDINAMARCA
FACULTAD DE INGENIERÍA
PROGRAMA INGENIERÍA DE SISTEMAS
FACATATIVÁ, MAYO DE 2018**

DEDICATORIA

AGRADECIMIENTOS

COMPROMISO DEL AUTOR

Yo, Oscar Jobany Gómez Ochoa identificado con cédula de ciudadanía No. 1.073'427.694 de Bojacá – Cundinamarca y con código No. 461212116, estudiante del programa de Ingeniería de Sistemas de la Universidad de Cundinamarca, declaro que:

El contenido del presente documento es un reflejo de mi trabajo personal y manifiesto que, ante cualquier notificación de plagio, copia o falta a la fuente original, soy responsable directo legal, económico y administrativo sin afectar al director del trabajo, a la Universidad y a cuantas instituciones hayan colaborado en dicho trabajo, asumiendo las consecuencias derivadas de tales prácticas.

En constancia de lo anterior firma:



Oscar Jobany Gómez Ochoa

C.C. 1.073'427.694 Bojacá - Cundinamarca

RESUMEN

El presente documento muestra el desarrollo de un sistema inteligente para el registro e identificación de Lepidópteros – Rhopalocera que habitan en el parque Metropolitano María Lucía (PMML) en Villavicencio - Colombia, mediante el uso de Redes Neuronales Artificiales (RNA) y Modelos Ocultos de Markov (HMM). Inicialmente se registró información de cinco especies de mariposas y fotografías de cada una, a las cuales se les aplicó técnicas de procesamiento de imágenes para lograr la extracción de características e identificación de patrones. Esta información se adicionó a una base de datos local y se utilizó para entrenar el aplicativo, dotándolo de inteligencia. En la validación de la herramienta, se le presentaron imágenes diferentes, similares o iguales a las utilizadas en la fase de aprendizaje; la respuesta del programa se busca en la base de datos local, si existe coincidencia, muestra la información correspondiente de dicha especie, de lo contrario indica que era una especie desconocida, con ello y la experticia de un investigador del mariposario del PMML se determinó si las respuestas del sistema eran acertadas o se requería realizar el proceso de entrenamiento nuevamente. Esta herramienta fue propuesta y desarrollada por integrantes del Centro de Innovación y Tecnología (CIT) de la Universidad de Cundinamarca (UDEC) extensión Facatativá para brindar apoyo tecnológico al proyecto “Diseño e implementación de un mariposario” adscrito al Centro de Investigaciones Ambientales (CIAM) de la Corporación Universitaria de Meta (UNIMETA).

Palabras Clave: Lepidópteros, Modelo Oculto de Markov, procesamiento de imágenes, Red Neuronal Artificial.

ABSTRACT

This document shows the development of an intelligent system for the registration and identification of Lepidoptera - Rhopalocera that inhabit the María Lucía Metropolitan Park (PMML) in Villavicencio - Colombia, through the use of Artificial Neural Networks (RNA) and Hidden Markov Models (HMM) Initially information was recorded on five species of butterflies and photographs of each one, which applied image processing techniques to achieve the extraction of characteristics and the identification of patterns. This information is added to a local database and used for the use of the protocol, providing it with intelligence. In the validation of the tool, the different images, similar or equal to those used in the learning phase; the response of the program is searched in the local database, if there is a coincidence, the information associated with this species, otherwise it indicates that it was an unknown species, with this and the expert of a researcher of the PMML butterfly was determined if the System responses were accurate or required to complete the training process again. This tool was proposed and developed by members of the Center for Innovation and Technology (CIT) of the University of Cundinamarca (UDEC) Facatativa extension to provide technological support to the project "Design and implementation of a butterfly" attached to the Center for Environmental Research (CIAM) of the University Corporation of Meta (UNIMETA).

Key Words: *Artificial Neural Network, Hidden Markov Model, image processing, Lepidoptera.*

INDICE GENERAL

| | |
|--|-----|
| DEDICATORIA..... | IV |
| AGRADECIMIENTOS..... | V |
| COMPROMISO DEL AUTOR..... | VI |
| RESUMEN..... | VII |
| INDICE GENERAL..... | IX |
| LISTA DE TABLAS..... | XI |
| LISTA DE FIGURAS..... | XII |
| LISTA DE ANEXOS..... | XV |
| INTRODUCCIÓN..... | XVI |
| I. INFORME DE INVESTIGACIÓN..... | 1 |
| 1.1. ESTADO DEL ARTE..... | 1 |
| 1.2. LÍNEA DE INVESTIGACIÓN..... | 4 |
| 1.3. PLANTEAMIENTO DEL PROBLEMA..... | 5 |
| 1.4. OBJETIVO GENERAL Y OBJETIVOS ESPECÍFICOS..... | 6 |
| 1.4.1. OBJETIVO GENERAL..... | 6 |
| 1.4.2. OBJETIVOS ESPECÍFICOS..... | 6 |
| 1.5. ALCANCE E IMPACTO DEL PROYECTO..... | 6 |
| 1.6. METODOLOGÍA..... | 8 |
| 1.7. MARCOS DE REFERENCIA..... | 11 |
| 1.7.1. MARCO HISTÓRICO..... | 11 |
| 1.7.2. MARCO TEÓRICO..... | 20 |
| 1.7.3. MARCO LEGAL..... | 33 |
| II. DOCUMENTACIÓN DEL SOFTWARE..... | 35 |
| 2.1. PLAN DE PROYECTO..... | 35 |
| 2.2. DETERMINACION DE REQUERIMIENTOS..... | 35 |
| 2.3. ESPECIFICACIÓN DEL SISTEMA..... | 40 |
| 2.4. ESPECIFICACIÓN DEL DISEÑO..... | 41 |
| 2.4.1. MODELO DE ENTIDAD RELACIÓN (MER)..... | 41 |
| 2.4.2. DIAGRAMAS DE CASOS DE USO..... | 41 |

| | | |
|--------|---|----|
| 2.4.3. | DIAGRAMAS DE SECUENCIAS | 43 |
| 2.4.4. | DIAGRAMAS DE ACTIVIDADES | 46 |
| 2.4.5. | DIAGRAMA DE CLASES | 48 |
| 2.4.6. | DIAGRAMA DE DESPLIEGUE | 49 |
| III. | DOCUMENTACION DEL SISTEMA INTELIGENTE | 50 |
| 3.1. | PROCESAMIENTO DE IMÁGENES | 50 |
| 3.2. | ENTRENAMIENTO DEL SISTEMA | 61 |
| 3.3. | VALIDACIÓN DE RESULTADOS | 63 |
| IV. | ESTIMACION DE RECURSOS | 65 |
| V. | RESULTADOS | 66 |
| 5.1. | CAPTURAS DE PANTALLA REIDMAR v.1.0 | 68 |
| VI. | CONCLUSIONES Y RECOMENDACIONES | 76 |
| VII. | BIBLIOGRAFIA | 77 |

LISTA DE TABLAS

| | |
|---|----|
| Tabla 1. Semilleros del GISTFA | 4 |
| Tabla 2. Requerimiento Funcional 1 | 35 |
| Tabla 3. <i>Requerimiento Funcional 2</i> | 36 |
| Tabla 4. Requerimiento Funcional 3 | 36 |
| Tabla 5. Requerimiento Funcional 4 | 37 |
| Tabla 6. Requerimiento Funcional 5 | 37 |
| Tabla 7. Requerimiento Funcional 6 | 38 |
| Tabla 8. Requerimiento Funcional 7 | 38 |
| Tabla 9. Requerimiento Funcional 8 | 39 |
| Tabla 10. Requerimiento Funcional 9 | 39 |
| Tabla 11. Requerimiento Funcional 10..... | 40 |
| Tabla 12. Recursos del Proyecto..... | 65 |
| Tabla 13. Porcentajes de acierto RNA por especie de Mariposa | 66 |
| Tabla 14. Porcentajes de acierto HMM por especie de Mariposa | 67 |

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1. Metodología del proyecto. | 9 |
| Figura 2. Proceso de convolución..... | 22 |
| Figura 3. Resultado de aplicar convolución. | 23 |
| Figura 4. Máscaras más usadas. | 23 |
| Figura 5. Interpolación de nivel de gris basado en el concepto del vecino más cercano. | 25 |
| Figura 6. Modelo Entidad Relación Base de Datos | 41 |
| Figura 7. Casos de Uso | 42 |
| Figura 8. Diagrama de Secuencia Ingreso al Sistema | 43 |
| Figura 9. Diagrama de Secuencia Gestión Mariposa | 44 |
| Figura 10. Diagrama de Secuencia Entrenamiento Sistema..... | 44 |
| Figura 11. Diagrama de Secuencia Uso del Sistema Administrador | 45 |
| Figura 12. Diagrama de Secuencia Aplicación Usuario | 45 |
| Figura 13. Diagrama de Actividades Ingreso. | 46 |
| Figura 14. Diagrama de Actividades Mariposas..... | 47 |
| Figura 15. Diagrama de Actividades Entrenamiento. | 47 |
| Figura 16. Diagrama de Actividades Usuarios..... | 48 |
| Figura 17. Diagrama de Clases..... | 48 |
| Figura 18. Diagrama de Despliegue..... | 49 |
| Figura 19. Metodología del proyecto..... | 50 |
| Figura 20. Foto de mariposa <i>Methona Confusa</i> en el PMML..... | 51 |

| | |
|--|----|
| Figura 21. Foto de mariposa <i>Dione Juno</i> en el PMML. | 51 |
| Figura 22. Foto de mariposa <i>Callicore Pitheas</i> en el PMML. | 52 |
| Figura 23. Foto de mariposa <i>Mechanitis Polymnia</i> en el PMML. | 53 |
| Figura 24. Foto de mariposa <i>Colobura Dirce</i> en el PMML..... | 53 |
| Figura 25. Interfaz administración de mariposas REIDMAR v1.0. | 54 |
| Figura 26. Segunda fase metodológica. | 55 |
| Figura 27. Selector de imágenes REIDMAR v1.0. | 55 |
| Figura 28. Tercera fase metodológica..... | 56 |
| Figura 29. Diagrama detallado para el procesamiento de las imágenes. | 56 |
| Figura 30. Técnica de suavizado aplicada. | 57 |
| Figura 31. Técnica escala grises aplicada. | 57 |
| Figura 32. Técnica detección de bordes aplicada..... | 58 |
| Figura 33. Técnica de transformación aplicada..... | 58 |
| Figura 34. Técnica de binarización aplicada..... | 59 |
| Figura 35. Cuarta Fase Metodológica. | 59 |
| Figura 36. Diagrama reconocimiento de patrones. | 60 |
| Figura 37. Diagrama entrenamiento RNA. | 61 |
| Figura 38. Diagrama entrenamiento RNA. | 62 |
| Figura 39. Diagrama entrenamiento HMM. | 62 |
| Figura 40. Interfaz validación administrador REIDMAR v1.0. | 63 |
| Figura 41. Presentación REIDMAR v1.0. | 68 |
| Figura 42. Menú Principal REIDMAR v1.0. | 69 |

| | |
|---|----|
| Figura 43. Acerca de REIDMAR v1.0..... | 70 |
| Figura 44. Menú Usuario General REIDMAR v1.0..... | 71 |
| Figura 45. Login Administrador REIDMAR v1.0 | 72 |
| Figura 46. Menú Administrador REIDMAR v1.0 | 73 |
| Figura 47. Menú Mariposas REIDMAR v1.0 | 74 |
| Figura 48. Menú Validación del Administrador REIDMAR v1.0..... | 75 |

LISTA DE ANEXOS

- ✓ Carta Aval
- ✓ MINr08_V5
- ✓ Requerimientos IEEE
- ✓ Manual de Usuario
- ✓ Manual Técnico
- ✓ Derechos de Autor
- ✓ Artículo CICI 2016
- ✓ Artículo CIING 2016
- ✓ Artículo EACI 2017
- ✓ Artículo CIINDET 2018
- ✓ CvLAC del Autor
- ✓ Datos de Contacto

INTRODUCCIÓN

Colombia, a nivel mundial, es el segundo país con mayor biodiversidad de mariposas, con aproximadamente 3274 especies (Andrade, 2011). A pesar de esta riqueza natural, es poco lo que se conoce de sus ciclos de vida, información básica para encaminar medidas de conservación de los lepidópteros en especial en la Orinoquia, donde la abundancia natural de mariposas suele disminuir en la medida que hay mayor intervención humana en sitios como ciudades (Rubio M, Jiménez F, 2015), así mismo el crecimiento de las áreas de cultivos en particular monocultivos y la explotación de hidrocarburos, ha generado la pérdida y transformación de las coberturas vegetales originales (Minorta & Rangel, 2014), con efectos nocivos para los lepidópteros (Constantino, 1997), asociados con la transformación de la vegetación que genera cambios en la composición y en la riqueza de especies (Francesconi, Nair, Levey, Daniels, Cullen, 2013), obteniendo como consecuencia escasa o nula presencia de mariposas en su hábitat de conservación.

Teniendo en cuenta esto y en aras de aumentar el conocimiento sobre lepidópteros, CIAM de la UNIMETA identifica, registra y controla cinco especies de mariposas (*Methona Confusa-Dione Juno-Callicore Pitheas-Mechanitis Polymnia-Colobura Dirce*) que habitan el PMML sobre el proyecto titulado como “Plan de manejo para la cría de lepidópteros como alternativa para la protección de ecosistemas en el parque metropolitano María Lucía” a cargo de la Licenciada en C N-MSc Gestión Ambiental Luz Dary Pedraza Hernández.

Para apoyar este proyecto el CIT de la UDEC extensión Facatativá, desarrolló una herramienta de software utilizando técnicas relacionadas con el reconocimiento de patrones, las cuales representan elementos importantes dentro de los sistemas inteligentes (Ray Liu, 2002), en específico se tomó el procesamiento, segmentación y análisis de imágenes en conjunto con métodos de inteligencia artificial.

La herramienta fue desarrollada en lenguaje de programación JAVA bajo tres aspectos de interés, el primero corresponde a las técnicas utilizadas para el procesamiento de imágenes y extracción de características, el segundo se refiere a la implementación de métodos de inteligencia artificial para el reconocimiento de patrones, para este proyecto fueron utilizadas las Redes Neuronales Artificiales (RNA) ya que estas “son muy buenas clasificadoras de patrones debido a su capacidad de aprender patrones que no son linealmente separables y conceptos que se ocupan de la incertidumbre, el ruido y los eventos aleatorios” (Dębska & Guzowska-Świder, 2011) y los Modelos Ocultos de Markov (HMM) que según (Jurafsky & Martin, 2017) “son modelos de secuencias probabilísticas o clasificadores cuya tarea es asignar una etiqueta o clase a cada unidad en una secuencia calculando la distribución de probabilidad sobre posibles secuencias de etiquetas y eligen la mejor de estas secuencias.”

Inicialmente se registraron las cinco mariposas a identificar (*Methona Confusa* - *Dione Juno* - *Callicore Pitheas* - *Mechanitis Polymnia* - *Colobura Dirce*) con datos ecológicos, morfológicos y algunos hábitos de cada especie. Luego, se cargaron al sistema veinte fotografías de cada ejemplar en estudio, a las cuales se les aplicaron técnicas de procesamiento de imágenes (Suavizado - Escala de Grises - Detección de Bordes - Transformación - Binarización) para lograr la extracción de características e identificación de patrones, esta información se adicionó a una base de datos local y se usó para entrenar el sistema dotándolo de inteligencia. En la validación de la herramienta, se le presentaron imágenes diferentes, similares o iguales a las utilizadas en la fase de aprendizaje, la respuesta que devolvía el aplicativo se buscaba en la base de datos local, si existía coincidencia, mostraba la información correspondiente de dicha especie, de lo contrario indicaba que era una especie desconocida, con esto y la experticia de un investigador del mariposario del PMML se determinó si las respuestas del sistema eran acertadas o se requería realizar el proceso de entrenamiento nuevamente. Finalmente utilizando una comparativa basada en los resultados obtenidos de los dos métodos (RNA – HMM)

se determinó cuál de los dos respondió mejor para la identificación y reconocimiento de las mariposas.

I. INFORME DE INVESTIGACIÓN

1.1. ESTADO DEL ARTE

En la actualidad existen muchas aplicaciones para reconocimiento de imágenes mediante patrones, la técnica más utilizada son las RNA que cuentan con diversos métodos de aprendizaje como el perceptrón, (Górska-Horczyzak, Horczyzak, Guzek, Wojtasik-Kalinowska, & Wierzbicka, 2017) utilizan un dispositivo que por medio del procesamiento de imágenes y la implementación de un perceptrón multicapa determina si la carne de cerdo se encuentra en buen estado. Otro de los métodos de aprendizaje de las RNA son las Backpropagation, (Murru & Rossini, 2016) aclaran que para utilizar este método es muy importante la correcta inicialización de pesos, tarea que por lo general se hace aleatoriamente, ellos ejecutando el filtro de Kalman demostraron mejoras de rendimiento frente a la generación aleatoria, ahora bien, en cuanto a la implementación de estas redes tenemos a (Kouakou, Bagui, Agneroh, Soro, & Zoueu, 2016) que muestran el uso de una backpropagation para detectar tempranamente el virus del mosaico en pepinos analizando la huella dactilar de las hojas de pepino, (Fan et al., 2013) quien crea un sistema de visión por computadora para el análisis de perfil de la textura de alimentos identificando la dureza y gomosidad de estos sin la necesidad de masticarlos mediante una red backpropagation, demostrando que la culinaria también puede ser apoyada por las RNA o también está (Olawoyin, 2016) el cual presenta un sistema para predecir e identificar los principales tóxicos presentes en el microambiente y los suelos para poder tomar decisiones en un futuro cercano y evitar aumento de la contaminación.

Por otro lado, (Hameed, Karlik, & Salman, 2016) proponen una versión modificada de la red backpropagation para mejorar su comportamiento, acelerando el proceso de convergencia para acceder al estado estacionario óptimo y minimizar el error mejorando los patrones, esta versión fue comparada con otros algoritmos de aprendizaje supervisado dando excelentes resultados en cuanto a la velocidad de

convergencia. (Zhou, Zhang, Yin, & Liu, 2016) plantea un método de aprendizaje para el reconocimiento de caracteres chinos escritos a mano usando el aprendizaje de rasgos cuadráticos discriminatorios, el cual supera aplicaciones similares en más de un 4% en la precisión del reconocimiento y la velocidad con la que se reconocen los caracteres, mientras que, (Surinta, Karaaba, Schomaker, & Wiering, 2015) utilizan descriptores de rasgos de gradiente local que es usado para la extracción de vectores característicos de las imágenes manuscritas presentándolas luego al algoritmo de aprendizaje para realizar la respectiva clasificación que para este caso es evaluada en tres scripts de idioma Tailandés, Latino y Bangladesh, los resultados muestran que los descriptores de características de gradiente local realizan un estudio significativo de la intensidad de los píxeles en las imágenes que combinados con los receptores de características se obtienen altas precisiones en los conjuntos de datos escritos a mano pero (Mofavvaz, Sohrabi, & Nezamzadeh-Ejhi, 2017) hacen uso de una metodología llamada máquinas de vector de soporte de mínimos cuadrados con algoritmos de entrenamiento diferentes, realizando una comparación entre Levenberg-Marquardt - LM y gradiente descendente con momento y algoritmo de retro-proporción adaptativa - GDX de los cuales al realizaron las respectivas comparaciones sin encontrar significativas entre ellos.

El reconocimiento de la expresión facial se ha convertido en una obsesión desde hace unos 10 años y para ello, (Lopes, de Aguiar, De Souza, & Oliveira-Santos, 2017) se basan en redes neuronales artificiales convolucionales, que se ajustan mejor a la precisión de los datos, con este proyecto se logran resultados de competitividad comparados con otros métodos de reconocimiento de la expresión facial además de permitir la reconfiguración de la factibilidad de tiempo real.

En cuanto a las mariposas, (Kaya, Kayci, & Uyar, 2015) exponen un método de reconocimiento mediante redes neuronales artificiales que se basa en un patrón binario local para clasificarlas en función de sus cualidades morfológicas externas logrando con alto porcentaje de reconocimiento en termino de tasa de precisión para la correcta identificación de mariposas.

Revisando otros métodos de reconocimiento de patrones, se encuentran los modelos ocultos de Markov (HMM), presentando a (Lisberg et al., 2016) quien utilizando los HMM y mediante la variabilidad en el consumo de energía en los edificios, facilito identificar el comportamiento de sus ocupantes, creando métodos para observaciones indirectas y caracterización del comportamiento de los ocupantes teniendo en cuenta dependencias tales como temperatura del aire ambiente y encuestas de ocupantes las cuales son dependientes del tiempo para mejorar la predicción obteniendo buenos resultados para mitigar el consumo de energía del edificio. O también HMM implementados en un sistema automático para la detección de eventos volcánicos dado que estos tienen diferentes duraciones, por lo tanto, el estado y toda la duración de los eventos aprendidos de los datos de entrenamiento se aplican en el estado correspondiente y los modelos de duración del evento dentro del HMM, (Bhatti et al., 2016) realizo los experimentos necesarios para demostrar que un HMM estándar puede detectar los eventos volcánicos con alta precisión, pero genera falsos positivos. Y combinando los HMM y el reconocimiento de gestos de las manos llega (Premaratne, Yang, Vial, & Ifthikar, 2017) con su propuesta de la introducción de seguimiento centroide de los gestos de las manos para superar las limitaciones del inadecuado uso de la información de tiempo, del uso a gestos simples y datos subutilizados.

Una imagen monocromática, tradicionalmente se presenta como una matriz rectangular cuyos valores describen nivel de intensidad o luminancia conocido como pixel, esta representación es la más utilizada para procesamiento de imágenes pero (Jimenez-Fernandez et al., 2016) trae una nueva alternativa que se trata de utilizar funciones continuas por partes y lineales de dos variables espaciales como una descripción del modelo monocromático en una imagen, mejorando la calidad de la misma con respecto al método tradicional y abriendo las puertas a la implementación de funciones más robustas.

1.2. LÍNEA DE INVESTIGACIÓN

El Grupo de Investigación de Sistemas y Tecnología de Facatativá (GISTFA), creado en el año 2007 y con código de grupo certificado COLCIENCIAS No. COL0085019, hace parte del programa de Ingeniería de Sistemas de la Universidad de Cundinamarca (UDEC) Extension Facatativá. El GISTFA se alimenta de los semilleros LINUDEX, CLUOD-IN, SERFAC y APLICACIONES MOVILES, en la tabla 1 se muestra el docente líder de cada semillero activo a la fecha.

Lista de Semilleros GITSFA

| Nombre Semillero | Líder Semillero |
|-------------------------|----------------------------|
| LINUDEX | Gina Maribel Valenzuela S. |
| CLOUD-IN | Oscar Javier Morera Z. |
| SERFAC | Jaime Eduardo Andrade R. |
| APLICACIONES MOVILES | Cesar Yesid Barahona R. |

Tabla 1. Semilleros del GISTFA

GISTFA trabaja bajo la línea de investigación **Software, sistemas emergentes y nuevas tecnologías**, definida por la UDEC Extension Facatativá. A la fecha, el grupo cuenta con cinco proyectos de investigación avalados por la UDEC y seis investigadores reconocidos ante COLCIENCIAS.

Es así como GISTFA asume la responsabilidad de la investigación del programa con la tarea de convertir los resultados de sus semilleros en proyectos eficientes, óptimos y de alto impacto en la comunidad académica y su entorno. Sobre esta base y bajo convenios hechos con las empresas de la región el GISTFA ha implementado proyectos de desarrollo e innovación en la Alcaldía Municipal, Escuela de Comunicaciones, Policía Nacional y la empresa de aguas de Facatativá. Fue relevante para el grupo los trabajos desarrollados con el Instituto de Desarrollo,

Recreación y Deporte (IDRD) del Distrito Capital entre otras. Otros intereses de la investigación, en el GISTFA los comprometen hacia una labor en la cual un ejercicio permanente actualice su razón de ser a partir de proyectos informáticos que generen conocimiento pertinente y consistente en el tiempo. Ahora bien, el GISTFA crece y así las investigaciones orientadas por el grupo, han contribuido a la creación de ambiciosas metas como la generación de proyectos interdisciplinarios (en el caso de la alianza con el grupo e investigación AXIOMA) y la transición hacia el emprendimiento y la innovación para fortalecer la transferencia de conocimiento y las conexiones con el sector empresarias.

Actualmente y desde su labor investigativa, el GISTFA cuenta con el Centro de Investigaciones y Tecnologías (CIT) en el programa de Ingeniería de Sistemas UDEC Extensión Facatativá, con el objetivo de aumentar el impacto de los productos generados por el GISTFA en la evaluación realizada por COLCIENCIAS y así mismo proponer nuevas metodologías de desarrollo como resultado de los productos generados a lo largo de la vida del proyecto informático.

1.3. PLANTEAMIENTO DEL PROBLEMA

El grupo GISTFA evidenció, en un primer acercamiento con la UNIMETA, que el grupo CIAM registra e identifica de forma manual cinco especies de mariposas que habitan el parque María Lucía, sabiendo esto se consideró que sería útil tener una herramienta tecnológica que facilitara dicha tarea.

¿Puede un sistema inteligente dar apoyo tecnológico al proceso de registro e identificación de mariposas en el parque Metropolitano María Lucía en Villavicencio que realiza el CIAM de la UNIMETA?

1.4. OBJETIVO GENERAL Y OBJETIVOS ESPECÍFICOS

1.4.1. OBJETIVO GENERAL

Desarrollar un prototipo de sistema inteligente como herramienta tecnológica para el registro e identificación de cinco especies de mariposas que habitan el parque metropolitano María Lucia en Villavicencio-Meta.

1.4.2. OBJETIVOS ESPECÍFICOS

1. Realizar modelado y diseño del sistema inteligente basados en los requerimientos de la Corporación Universitaria del Meta.
2. Desarrollar los módulos de procesamiento de imágenes, almacenamiento, aprendizaje/aplicación y comparativa del sistema.
3. Hacer pruebas de funcionalidad del sistema inteligente en el parque metropolitano María Lucia en Villavicencio.
4. Establecer, por medio de los resultados obtenidos, cuál de los métodos de reconocimiento de patrones implementados en el sistema es el más adecuado para el proceso que realiza el CIAM con las mariposas.

1.5. ALCANCE E IMPACTO DEL PROYECTO

Con este proyecto se buscó apoyar tecnológicamente el proceso de registro e identificación de las cinco especies de mariposas que habitan el Parque Metropolitano María Lucia ubicado en la ciudad de Villavicencio – Meta. El CIAM de la UNIMETA contará con una herramienta útil y de fácil manejo para el registro e identificación de las mariposas del parque, ya que según la Lic. Luz Dary Pedraza: “tecnificar este proceso ayudaría para la certificación y licencia de exportación de

las mariposas”. Además se pretende formar una alianza, no solo interdisciplinaria entre los programas Ingeniería de Sistemas e Ingeniería Ambiental sino interinstitucional entre la Universidad de Cundinamarca y la Corporación Universitaria del Meta; esto conllevará a fortalecer los grupos de investigación de las dos instituciones y abrir las puertas a más proyectos de este tipo.

La idea de este proyecto nació en el año 2016 y desde entonces hasta la fecha se cuenta con los siguientes productos:

- Ponencia en el Congreso Internacional de Ciencias Básicas e Ingeniería (CICI) 2016 – Villavicencio, Colombia. Con el artículo titulado “Prototipo de Sistema Inteligente para el Registro e Identificación de Mariposas Nativas en la Región Orinoquia utilizando Redes Neuronales Artificiales implementadas en Java.” Con número ISBN: 978-958-8927-23-7.
- Registro de Software ante la Dirección Nacional de Derechos de Autor con No. De Certificado 13-59-238 del 29 de marzo del 2017.
- Llamado a ponencia en el XII Congreso Internacional de “Electrónica y Tecnologías de Avanzada” (CIETA) 2017 – Pamplona, Norte de Santander, Colombia. Con el artículo titulado “Prototipo de Sistema Inteligente para el Registro e Identificación de Mariposas Nativas del Parque María Lucía (Villavicencio) mediante Modelos Ocultos de Markov (HMM) Implementados en Java”.
- Ponencia en el Primer Encuentro Académico de Creatividad e Innovación (EACI) 2017 – Facatativá, Colombia. Con el artículo titulado “Inteligencia Artificial como Herramienta para Reconocimiento de Mariposas en Villavicencio (Caso de Estudio Parque Metropolitano María Lucía”.
- Llamado a ponencia en el VI Congreso de Ingeniería, Ciencia y Tecnología (ESTEC) 2107 – Ciudad de Panamá, Panamá. Con el artículo titulado “Sistema Inteligente para Registro e Identificación de Mariposas Habitantes del Parque Metropolitano María Lucía en Villavicencio haciendo uso de Redes Neuronales Artificiales”.

- Ponencia en el VI Congreso Internacional de Ingeniería (CIING) 2017 – Fusagasugá, Colombia. Con el artículo titulado “Prototipo de sistema inteligente para la identificación y registro de mariposas nativas de la región Orinoquia mediante el uso de redes neuronales artificiales implementadas en Java”.
- Ponencia en el XIV Congreso Internacional sobre Innovación y Desarrollo Tecnológico (CIINDET) 2018 – Cuernavaca, Morelos, México. Con el artículo titulado “Registro e Identificación de Mariposas en Villavicencio-Colombia Aplicando Inteligencia Artificial (Caso de Estudio Parque Metropolitano María Lucía)”. Con número ISBN: 978-607-95255-8-3.

1.6. METODOLOGÍA

Para el desarrollo del sistema se implementó parcialmente la metodología ágil Scrum que es una de las más populares en la actualidad por ser adaptable, iterativo, flexible y eficaz garantizando la transparencia en la comunicación y generando responsabilidad tanto individual como colectiva de procesos continuos (SCRUM Study, 2017). El despliegue del proyecto incluye el proceso completo desde el registro de las mariposas hasta la validación de los dos métodos inteligentes (RNA – HMM) para la identificación de las especies en estudio. En la **Figura 1** se muestran las fases de la metodología aplicada a la ejecución de este proyecto.

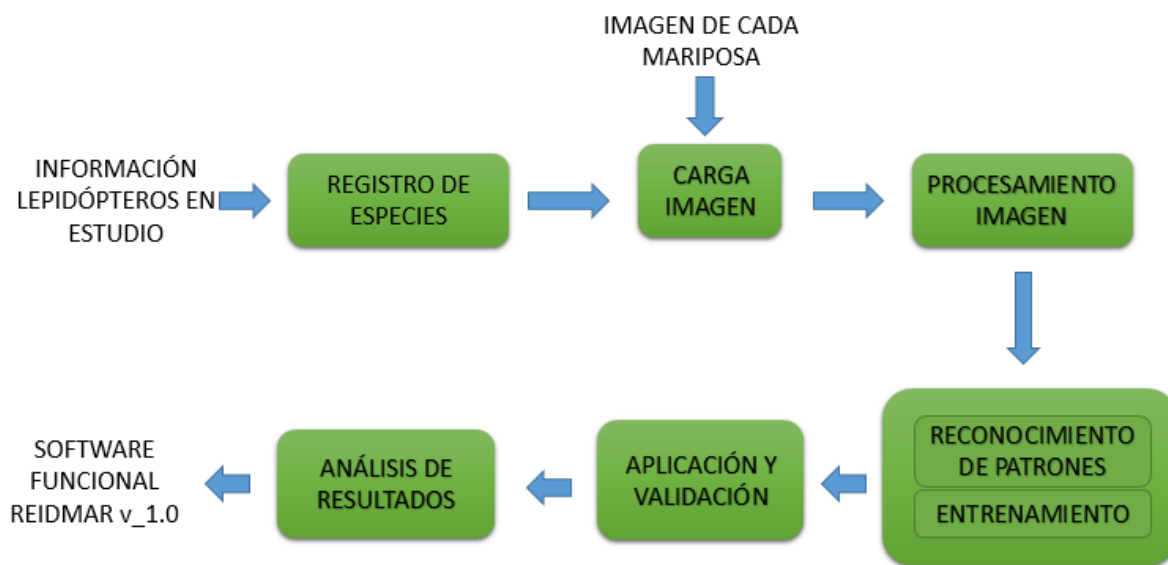


Figura 1. Metodología del proyecto.

En la primera fase se registraron las cinco especies de mariposas a identificar (*Methona Confusa* - *Dione Juno* - *Callicore Pitheas* - *Mechanitis Polymnia* - *Colobura Dirce*), estas especies se encuentran actualmente en estudio activo con características ecológicas, morfológicas y de comportamiento ya identificadas por el grupo de investigación CIAM en el PMML. La segunda fase correspondió a la digitalización de veinte fotografías por cada ejemplar en estudio, pasándolas de la cámara digital Samsung ES95, utilizada para realizar las tomas en el mariposario del PMML, al computador y luego cargándolas al sistema mediante un selector de archivos que sólo acepta formatos de imagen JPEG/JPG y PNG con las propiedades de tamaño por defecto de la cámara. La tercera fase se ejecutó con la aplicación de técnicas de procesamiento de imágenes como el suavizado que se utilizó para reducción del ruido, la escala de grises que convirtió la característica del color de las imágenes a formato monocromático (Blanco/Negro), la detección de bordes que buscó resaltar los cambios bruscos en el color de cada imagen, la transformación que se usó para ajustar y estandarizar el tamaño de las fotografías dentro del sistema y la binarización que permitió la distinción de características para

cada especie e identificación de patrones. En la cuarta fase, por medio de operaciones matemáticas, se convirtió cada imagen de mariposa procesada en un vector característico único de la especie y junto con el resto de los patrones u observaciones se inició el entrenamiento tanto de la RNA como del HMM. Para la quinta fase se desarrolló un módulo de aplicación, que se ejecuta después de terminado el entrenamiento, el cual consistió en presentarle imágenes iguales, similares o diferentes a las utilizadas en la fase de aprendizaje obteniendo una respuesta del sistema y con ayuda de la experticia de un investigador del PMML se validó si la respuesta del sistema era correcta o se hacía necesario un nuevo entrenamiento de la herramienta. Finalmente, utilizando los registros de resultados provenientes de la fase anterior se determinó cuál de los métodos inteligentes en cuestión (RNA - HMM) responde eficientemente frente a la identificación de patrones y reconocimiento de las mariposas en estudio.

1.7. MARCOS DE REFERENCIA

1.7.1. MARCO HISTÓRICO

Como ya se mencionó, este proyecto aborda temas como procesamiento de imágenes, red neuronal artificial, reconocimiento de patrones, modelos ocultos de markov y desarrollo de software. Como ya se sabe, ningún proyecto comienza de cero, es decir, tiene antecedentes que son la base para su desarrollo; a continuación mostraremos los antecedentes tenidos en cuenta en la investigación de nuestro proyecto:

1970 “Técnicas en el Procesamiento Digital de Imágenes”: El procesamiento digital de imágenes es el resultado de la interacción de varias áreas de estudio –como el análisis matemático y el tratamiento de señales– guiadas por diferentes técnicas de programación y optimización. Este artículo busca mostrar, a aquellas personas que deseen comenzar a trabajar con imágenes, un modelo matemático sencillo que permita visualizarlas como señales bidimensionales y entender su comportamiento de una forma básica al ser procesadas por medios digitales. De la misma forma, exponer algunas de las nuevas tendencias que se trabajan en la actualidad sobre este tema. El principal objetivo del procesamiento digital de imágenes es proponer y visualizar la viabilidad de diferentes soluciones a un problema específico, aplicándolas sobre una función en un espacio bidimensional (imágenes) o, en algunos casos, tridimensional (secuencias de imágenes). Usualmente, encontrar una solución óptima implica el planteamiento de un algoritmo; éste, se basa en una teoría que debe ser comprobada haciendo un extenso trabajo experimental a través de un *software* de simulación y grandes cantidades de imágenes de prueba (Melo, 1970).

1990 “Hidden Markov Models for Speech Recognition”: El uso de modelos de Markov ocultos para el reconocimiento del habla se ha convertido en predominante en los últimos años, como lo demuestra el número de artículos publicados y charlas en las principales conferencias de habla. Las razones por las que este método se ha vuelto tan popular son el marco estadístico inherente (matemáticamente preciso): la facilidad y disponibilidad de algoritmos de entrenamiento para estimar los parámetros de los modelos a partir de conjuntos de entrenamiento finito de datos de voz; la flexibilidad del sistema de reconocimiento resultante en el que se puede cambiar fácilmente el tamaño, tipo o arquitectura de los modelos para adaptarse a palabras, sonidos particulares. Etcétera; y la facilidad de implementación del sistema general de reconocimiento. En este artículo expositivo abordamos el papel de los métodos estadísticos en esta poderosa tecnología aplicada al reconocimiento del habla y discutimos una serie de cuestiones teóricas y prácticas que aún no han sido resueltas en términos de su importancia y su efecto en el rendimiento para las diferentes implementaciones del sistema (Huang, Jack, & Ariki, 1990).

2002 “Reconocimiento de Formas Manuscritas con Modelos Ocultos de Markov”: Desde que el hombre es capaz de usar su inteligencia, materializa sus ideas en palabras o expresiones, que expresa mediante su aparato de fonación y representa mediante signos gráficos. En la actualidad, el proceso de escritura y lectura implica el análisis de la serie de signos gráficos, lo que permite reconstruir su aspecto auditivo y visual, análisis que en su conjunto sugiere el concepto o la idea. El problema del reconocimiento de formas es variable según el planteamiento propuesto. Pero, ¿Se podría crear una técnica universal, que construya un operador para reconocer formas? Esto se creía en los años cincuenta, en los que se veían florecer propuestas de sistemas auto organizadores y de aprendizajes automáticos basado en perceptrones. En realidad, no existe ninguna aplicación capaz de construir un operador que resuelva cualquier problema del reconocimiento de formas. Por eso, se recomienda indagar en la modestia y en la paciencia, ya que en la actualidad, estamos muy lejos de hacerlo tan bien con nuestras potentes máquinas "de quinta generación"; como lo hacen los seres vivos con sus sentidos

del olor y de la vista. Por tanto, se nos presenta un campo amplio y en plena expansión, como se demuestra con las miles de complejas publicaciones por año. Este reconocimiento es un acercamiento de las distintas técnicas de la programación y la matemática, y no por la síntesis del algoritmo o de la teoría, sino por todas las formas que nos podemos encontrar en nuestro entorno más cercano (Manuel & González, 2002).

2004 “Procesamiento y análisis de imágenes digitales”: Una imagen natural capturada con una cámara, un telescopio, un microscopio o cualquier otro tipo de instrumento óptico presenta una variación de sombras y tonos. Imágenes de este tipo se llaman imágenes analógicas. Para que una imagen analógica, en blanco y negro, en escala de grises o a color puede ser "manipulada" usando un ordenador, primero deben convertirse a un formato adecuado. Este formato es la imagen digital correspondiente. La transformación de una imagen analógica a otra discreta se llama digitalización (de las coordenadas espaciales x , y) y es el primer paso en cualquier aplicación de procesamiento de imágenes digitales (Hoyos & Hoyos, 2004).

2007 “Reconocimiento de Patrones en un Arreglo Sensorico Usando Redes Neuronales”: En este trabajo se presenta el desarrollo de un sistema para el procesamiento de señales provenientes de un arreglo Sensorico mediante la implementación de redes neuronales artificiales. Ejemplos simples de aplicación se presentan para demostrar la potencialidad del sistema y su adaptabilidad a diferentes entornos (Rojas, Sarmiento, Lawrence, & Escobar, 2007).

2008 “Probalistic Principal Component Analysis Based Hidden Markov Model for Audio – Visual Speech Recognition”: Lipreading es un método eficiente entre los que se proponen para mejorar el rendimiento de los sistemas de reconocimiento de voz, especialmente en entornos acústicos ruidosos. Este trabajo propone un sistema de reconocimiento de voz audiovisual (AVSR), que podría mejorar la

robustez y precisión del reconocimiento de voz de audio integrando la información de audio y visual sincrónica. Proponemos un modelo de Markov oculto (HMM) basado en el análisis de componentes principales probabilístico (PCA) para el reconocimiento de voz sólo visual y la modalidad visual del reconocimiento de voz audiovisual. La HMM probabilística basada en PCA utiliza directamente las imágenes que sólo contienen la región de la boca del orador sin pre procesamiento (detección de la esquina de la boca, marcado de contorno, etc.) y toma PCA probabilística como la función de densidad de probabilidad de observación. Luego integramos estas dos modalidades de información (audio y visual) juntas y obtenemos un modelo de Markov oculto de varias corrientes (MSHMM). Encontramos que, sin extraer las características especializadas antes del procesamiento, el PCA probabilístico podría capturar los componentes principales durante el entrenamiento y describir la parte visual de los materiales. También se verifica por los experimentos que la integración de la información visual y de audio podría ayudar a mejorar la precisión de reconocimiento incluso a una baja relación señal ruido acústico (SNR) (Zhanyu Ma and Arne Leijon, 2008).

2009 “Reconocimiento de Imágenes a través de su Contenido”, Antonio Saenz: Los buscadores de imágenes actuales, como *Google images*, se basan principalmente en el nombre del archivo que contiene la imagen, en el nombre del enlace y en el texto que aparece en la página donde se encuentra la imagen. Es decir que la búsqueda se basa en información de tipo texto y no en información gráfica. El objetivo de este proyecto es desarrollar algoritmos de tratamiento digital de imágenes que permitan clasificar fotografías en base al contenido gráfico. La idea para el desarrollo de este proyecto se basa en la creación de un conjunto de imágenes para posteriormente segmentar la imagen, y extraer de ella objetos y fondos conocidos, y clasificarlas de acuerdo a los elementos localizados (Sáez, 2009).

2010 “Técnicas de Procesamiento Digital de Imágenes: Una Aplicación para Identificación de Personas a Través de los Senos Frontales”, Molder, Pastore: Este trabajo presenta un método de comparación de senos frontales, utilizando técnicas de Procesamiento Digital de Imágenes. La identificación de una persona a partir de la comparación de los senos frontales de su cráneo es admitida como un método válido de identificación positiva: la radiografía pre-mortem se compara con la radiografía post-mortem. Sin embargo, en virtud de la posición y distancia con que las radiografías han sido adquiridas esta comparación no puede ser realizada con una simple inspección visual. Las técnicas de Procesamiento Digital de Imágenes se constituyen en un importante aporte en esta disciplina ya que proporcionan una descripción cuantitativa de curvas cerradas a partir de invariantes. En este trabajo se presenta un método de comparación a partir de una selección de Factores de formas, invariantes al escalamiento, rotación y traslación. Se trabajó con imágenes reales provenientes de radiografías digitales y simuladas por software. Se obtuvo un parámetro de distancia mínima entre curvas que indica la similitud entre dos formas (Moler & Pastore, 2010).

2011 “Estado del Conocimiento de la Biodiversidad en Colombia y sus Amenazas”

Andrade: En este trabajo se evalúa el estado del conocimiento, endemismos, y especies amenazadas de la biota colombiana, con un análisis de la información disponible a la fecha, estudiando el estado de conservación de los principales ecosistemas como son los humedales, páramos, y áreas protegidas comparándolos a la luz de problemas como el cambio climático, la deforestación y minería (Andrade C, 2011).

2011 “Métodos de Umbralización de Imágenes Digitales”, Carlos Cattaneo:

La umbralización de imágenes es el proceso por el cual se busca un umbral óptimo que permita distinguir en una imagen los objetos del fondo de los objetos del primer plano. Este umbral es el punto (o valor) en el cual el histograma de una imagen se divide en dos picos. En la mayoría de las imágenes este valor resulta un poco difícil

de encontrar gráficamente debido a la complejidad de estos histogramas. Es por eso que se usan métodos paramétricos y no paramétricos que modelizan el problema y encuentran diferentes maneras de obtener este umbral. Pun y Kapur utilizaron la entropía de Shannon, haciendo sus modificaciones de acuerdo a su análisis y plantearon sus propuestas en la búsqueda de un umbral. Yen definió un método basado en entropía de la correlación de acuerdo a los lineamientos de Shannon. Kittler e Illingworth, propusieron un método basado en clusterización de grupos mediante el mínimo error del promedio de clasificación de píxeles. Continuando un trabajo anterior se agregan estos métodos en el cálculo del umbral en imágenes en gris, realizando luego la Binarización por capas, y comparando cada una de estas binarizaciones con una realizada en escala de grises.

Experimentalmente se encuentra que las imágenes por umbralización siguiendo el método de Pun (a partir de aquellas en escala de grises) dan como resultado imágenes poco definidas, siendo aquellas obtenidas por los métodos de Kapur y Yen más definidas en detalles. A la vez, el método propuesto por Kittler brinda resultados muy similares a Yen, umbralización muy parejas. Tanto Yen como Kittler tienen un comportamiento sobresaliente respecto de los otros métodos (Cattaneo, Larcher, Ruggeri, Herrera, & Biasoni, 2011).

2013 “Binarización de Imágenes Digitales”, Rafael Magro: En este trabajo se exponen y analizan los métodos más conocidos para la Binarización y umbralización de imágenes digitales. Se incluyen algoritmos y pseudocódigo para la aplicación de filtros gráficos especialmente indicados para su empleo sobre fotografías en escala de grises. Se explica el proceso más sencillo y adecuado para su uso con el fin de agilizar el dibujado manual de ilustraciones científicas por medio de puntos. Los sistemas que se detallan simplifican y acortan notablemente el tiempo de producción de dichas ilustraciones. Se hace un recorrido sobre el panorama de los aplicativos que existen en el mercado con capacidad para Binarización y tramado de imágenes digitales (Im et al., 2013).

2014 “Reconocimiento de Gestos de Manos como Mecanismo de Interacción Humano-Computador”, Carolina García: La detección de gestos realizados por usuarios se está convirtiendo en uno de los más importantes mecanismos para la interacción humano-computador, debido a la naturalidad e intuitividad que ofrece, además de ser un mecanismo llamativo, lo que motiva a las personas a usarlo. Sin embargo, para realizar una dirección apropiada de gestos se está utilizando hardware especializado, el cual puede ser de difícil acceso por costos e infraestructura. En este trabajo se propone una estrategia de detección de gestos usando cámaras web (las cuales son hardware no especializado de fácil acceso en un computador estándar), donde a través de pre- procesamiento de imágenes se disminuye el ruido, y utilizando clasificadores como Maquinas de Soporte Vectorial, se hace la detección del gesto realizado por el usuario. Los resultados muestran que el desarrollo de un sistema de reconocimiento de gestos no es una tarea fácil debido a la anuencia de diversas condiciones al momento de la ejecución del sistema. Asimismo, es fundamental una adecuada localización de la región de interés dado que esta incluye en la extracción y clasificación de características (García Cortés, 2014).

2015 “Automatic identification of butterfly species based on local binary patterns and artificial neural”, Kaya, Kayci, Uyar: Las mariposas se clasifican primero en función de sus cualidades morfológicas externas. Se requiere analizar los caracteres genitales de los mismos cuando no es posible clasificar según las cualidades morfológicas externas. Las características genitales de una mariposa se pueden determinar usando varias sustancias químicas y métodos. Actualmente, estos procesos se realizan manualmente preparando diapositivas genitales de la mariposa recolectada a través de ciertos procesos determinados. Para algunos grupos de mariposas se deben aplicar técnicas moleculares para la identificación que es costoso de usar. En este estudio, se propone un método de visión por computadora para identificar automáticamente las especies de mariposas como una alternativa a los métodos de identificación convencionales. El método se basa en el

Patrón Binario Local (LBP) y la Red Neuronal Artificial (RNA). Un total de 50 imágenes de mariposas de cinco especies se utilizaron para evaluar la eficacia del método propuesto. Los resultados experimentales demostraron que el método propuesto ha logrado un buen reconocimiento en términos de tasas de precisión para la identificación de especies de mariposas (Kaya et al., 2015).

2016 “Prototipo De Sistema Inteligente Para La Identificación Y Registro De Mariposas Nativas De La Región Orinoquia Mediante Redes Neuronales Artificiales Implementadas En Java”, Gomez, Rodríguez, Andrade: En el presente documento se muestra el desarrollo de un prototipo de sistema inteligente para el reconocimiento y registro de información de mariposas que habitan la región Orinoquia colombiana y posterior identificación mediante el uso de redes neuronales artificiales (RNA) implementadas en Java. El sistema captura una fotografía del ejemplar en estudio, posteriormente se binariza la información y se aplican técnicas de procesamiento de imágenes hasta llegar a la extracción de características y la identificación de patrones. Seguidamente se comparan dichos resultados con los almacenados en una base de datos local y de ser iguales se presenta la información de dicha especie, en caso contrario, datos correspondientes al nombre, familia a la que pertenece y tamaño del nuevo ejemplar de mariposa, son registrados en el sistema. Luego, esta información se toma como patrón de entrenamiento para posteriores identificaciones (Andrade, Gomez, & Rodríguez, n.d.).

2017 “Inteligencia Artificial Como Herramienta Para Reconocimiento De Mariposas En Villavicencio (Caso De Estudio Parque Metropolitano María Lucia)”, Gómez, Ramírez, Andrade: Este documento presenta la fase inicial de un sistema inteligente para el registro e identificación de Lepidópteros – Rhopalocera que habitan en el parque Metropolitano María Lucía (PMML) en Villavicencio mediante el uso de Redes Neuronales Artificiales (RNA) implementadas en Java y aplicadas a cinco especies de mariposas (Methona confusa - Dione junio - Colobura dirce - Callicore phiteas - Hamadryas feronia). El prototipo recibe una fotografía del ejemplar en

estudio, la convierte a un formato monocromático (Blanco-Negro) y aplicando técnicas de procesamiento de imágenes (Transformación- Escalado-Binarización) se logra la extracción de características e identificación de patrones. Seguidamente se comparan estos resultados con los almacenados en una base de datos local y de ser iguales se presenta la información de dicha especie, en caso contrario, datos ecológicos, características ecológicas, hábitos y fisiología del nuevo ejemplar de mariposa serán solicitados por el sistema para posteriores identificaciones. Esta herramienta fue desarrollada por integrantes del Centro de Innovación y Tecnología (CIT) de la Universidad de Cundinamarca (UDEC) extensión Facatativá para brindar apoyo tecnológico al proyecto “Diseño e implementación de un mariposario” adscrito al Centro de Investigaciones Ambientales (CIAM) de la Corporación Universitaria de Meta (UNIMETA) (Jaime et al., 2017).

2018 “Registro e identificación de Mariposas en Villavicencio –Colombia Aplicando Inteligencia Artificial (Caso de Estudio Parque Metropolitano María Lucía”, Gómez, Pedraza, Andrade: El presente documento muestra el desarrollo de un sistema inteligente para el registro e identificación de Lepidópteros – Rhopalocera que habitan en el parque Metropolitano María Lucía (PMML) en Villavicencio, mediante el uso de Redes Neuronales Artificiales (RNA) implementadas en Java y aplicadas a cinco especies de mariposas. El sistema recibe una fotografía del ejemplar en estudio, aplica técnicas de procesamiento de imágenes (Suavizado-Detección de Bordes-Escalado-Binarización) y la convierte a un formato monocromático (Blanco-Negro) logrando la extracción de características e identificación de patrones. Seguidamente se comparan estos resultados con los almacenados en una base de datos y de ser iguales se presenta la información de dicha especie, en caso contrario, datos ecológicos, hábitos y fisiología del nuevo ejemplar de mariposa serán solicitados por el sistema para posteriores identificaciones. Finalmente, se valida el sistema con pruebas de campo en el mariposario del parque comparando los resultados del sistema con la opinión y criterios del experto. Esta herramienta fue propuesta y desarrollada por integrantes del Centro de Innovación y Tecnología

(CIT) de la Universidad de Cundinamarca (UDEC) extensión Facatativá para brindar apoyo tecnológico al proyecto “Diseño e implementación de un mariposario” adscrito al Centro de Investigaciones Ambientales (CIAM) de la Corporación Universitaria de Meta (UNIMETA).

1.7.2. MARCO TEÓRICO

En esta sección se explican los aspectos teóricos que abarcaron el desarrollo del sistema inteligente, el cual comprende temas relacionados con el procesamiento de imágenes, reconocimiento de patrones, inteligencia artificial y almacenamiento en base de datos.

Las técnicas relacionadas con el Reconocimiento de Patrones se consideran un elemento importante dentro de los sistemas inteligentes, principalmente de aquellos orientados al procesamiento de datos y toma de decisiones. El reconocimiento de patrones es un proceso utilizado para categorizar muestras de mediciones o datos observados como miembros de una de muchas categorías (SING-TZE, 2002). El reconocimiento de patrones es una tecnología en desarrollo acaparando el interés y la participación de muchas disciplinas, con lo cual, se puede encontrar para un mismo problema distintos enfoques y soluciones, compartiendo aspectos de interés como conjuntos difusos, inteligencia artificial, modelos neuronales, teoría de autómatas y aplicaciones en pre procesamiento, segmentación y análisis de imágenes, análisis sísmico, diagnósticos médicos, reconocimientos de rostros, voz, huella dactilar, entre otros (Seijas, 2001).

Respecto al procesamiento, segmentación y análisis de imágenes, existen diferentes medios para obtener una imagen digital, los más comunes son los scanner de cama plana y las cámaras digitales. Ambos se basan en un Dispositivo de Carga Acoplado o en inglés Coupled Charge Device (CCD), esta unidad recibe la luz de la imagen, ya sea por reflexión o por transmisión e integra en un tiempo

definido la cantidad de luz. El sistema más simple de digitalización corresponde al llamado tono de gris, este indica la cantidad o intensidad de luz registrada en un espacio corto de tiempo, por ejemplo el sistema utilizado en radiología. El principio de la cámaras digitales es la reflexión de la luz, solo que la fuente es externa a la cámara pero para todos los casos los datos capturados sufren un proceso de discretización o cuantización (Marín Ortiz, 2013). Estas imágenes capturadas con una cámara digital o cualquier otro tipo de instrumento óptico se conocen como imágenes naturales o analógicas ya sean en blanco y negro, en escala de grises o a color y para ser “manipuladas” en un ordenador primero deben convertirse a un formato discreto por medio de la digitalización que es el primer paso en cualquier aplicación de procesamiento de imágenes (Hoyos Hernández, 2004). Ya con la imagen digitalizada se utilizan técnicas para mejorar la calidad de la imagen disminuyendo el ruido, intensificando colores, suavizando bordes, conversión a escala de grises, binarización, entre otras, con el fin de extraer la mayor cantidad de información relevante (Hernández Nelson, 2004).

Una imagen digital puede interpretarse como una función bidimensional $z = F(x, y)$ donde x e y son coordenadas espaciales y z representa el valor de la intensidad en el punto (x, y) . En escala de grises, una imagen digital viene dada a partir de la matriz $z_{ij} = F(x_i, y_j)$, cada entrada de la matriz es un pixel y el número de pixeles (en el orden de la matriz) define la resolución de la imagen. El valor de cada pixel está cuantificado en un número determinado de bits desde 0 hasta 2^{p-1} posibles. Habitualmente se trabajan con $p = 8$ por lo que el rango de intensidades varía desde un valor 0 (negro) a 255 (blanco) pero cuando se trata de una imagen a color se debe trabajar con tres matices R, G y B que proporcionan las intensidades de los colores primarios rojo, verde y azul.

La técnica para filtrar imágenes más utilizada se llama convolución de matrices que consiste en la siguiente descripción: Dada una matriz $A_{m \times n}$ y una matriz $C_{(2N+1) \times (2N+1)}$ con $2N + 1 < m, n$ se establece el resultado de la convolución de las matrices A y C como una nueva matriz $D = A * C$ definida a partir de la expresión

$$d_{ij} = \frac{1}{c} \sum_{r=1}^{2N+1} \sum_{s=1}^{2n+1} a_{i-N+r-1, i-N+r-1} c_{r,s} \quad (1)$$

Donde $c = \sum_{i,j=1}^{2N+1} c_{ij}$ (si $c = 0$ se toma $c = 1$). Obsérvese que $d_{i,j}$ sólo está definido para $i = N + 1, \dots, m - N - 1$ y $j = N + 1, \dots, n - N - 1$.

La matriz C se le denomina núcleo o Kernel de la convolución.

Las dos siguientes figuras muestran el proceso de convolución de una matriz dada por otra de orden 3x3. La **Figura 2** recoge el resultado de la convolución correspondiente a las entradas (2,2) y (3,2). La **Figura 3** proporciona el resultado final de la convolución en donde a las entradas de la primera y última fila y primera y última columna se les ha asignado el valor original.

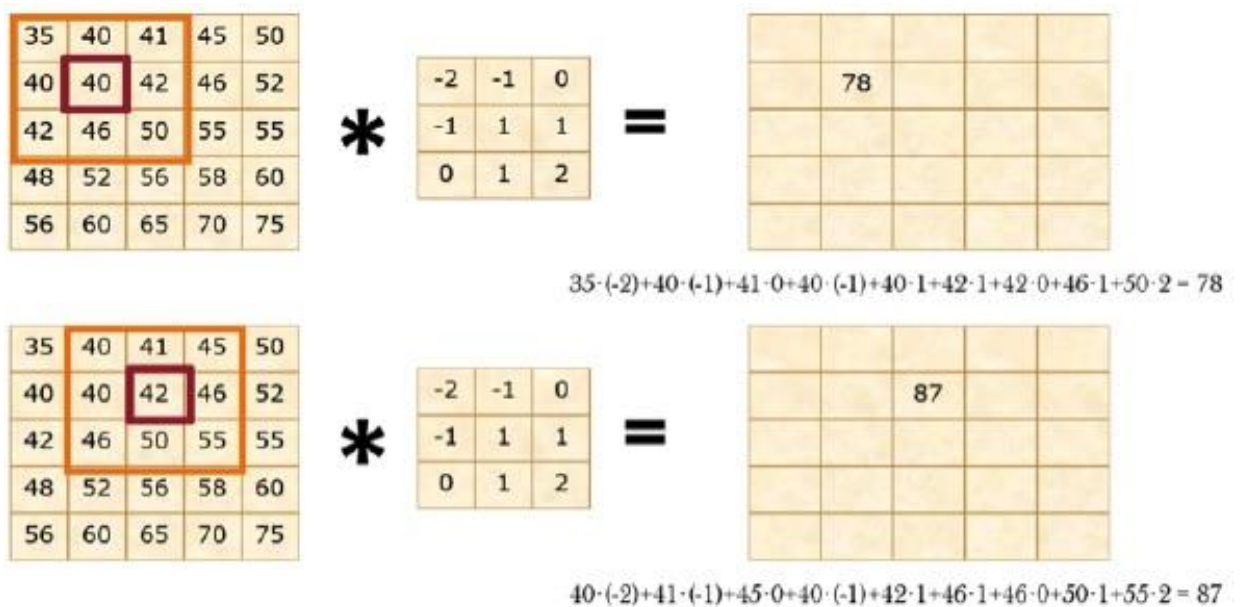


Figura 2. Proceso de convolución.

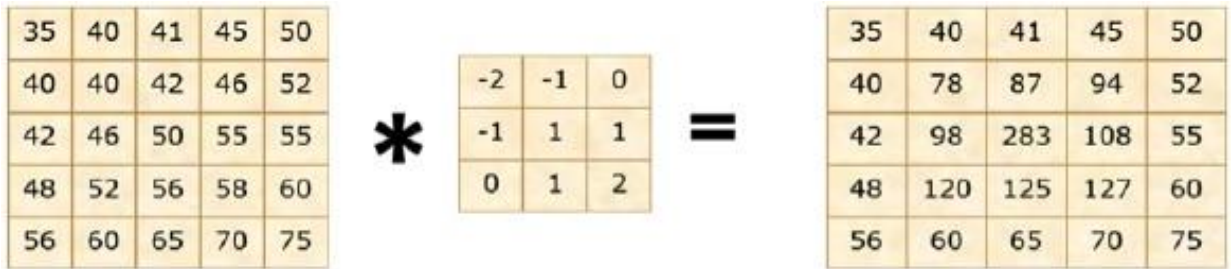


Figura 3. Resultado de aplicar convolución.

Para el filtrado de imágenes por lo general se usan matrices kernel o máscaras de orden 3x3 o 5x5. Para poder aplicar la convolución en los pixeles del borde de la imagen existen varias alternativas, dentro de las cuales se encuentra completar con ceros, repetir valores originales o completar la parte simétrica opuesta. Algunos de los filtros que se usan en el tratamiento de imágenes usan valores del parámetro c de la **Ecuación 1** arbitrarios, valor que suele ser usado para modificar el resultado final añadiendo un valor fijo a cada entrada. La **Figura 4** muestra algunas de las matrices máscaras más usadas.

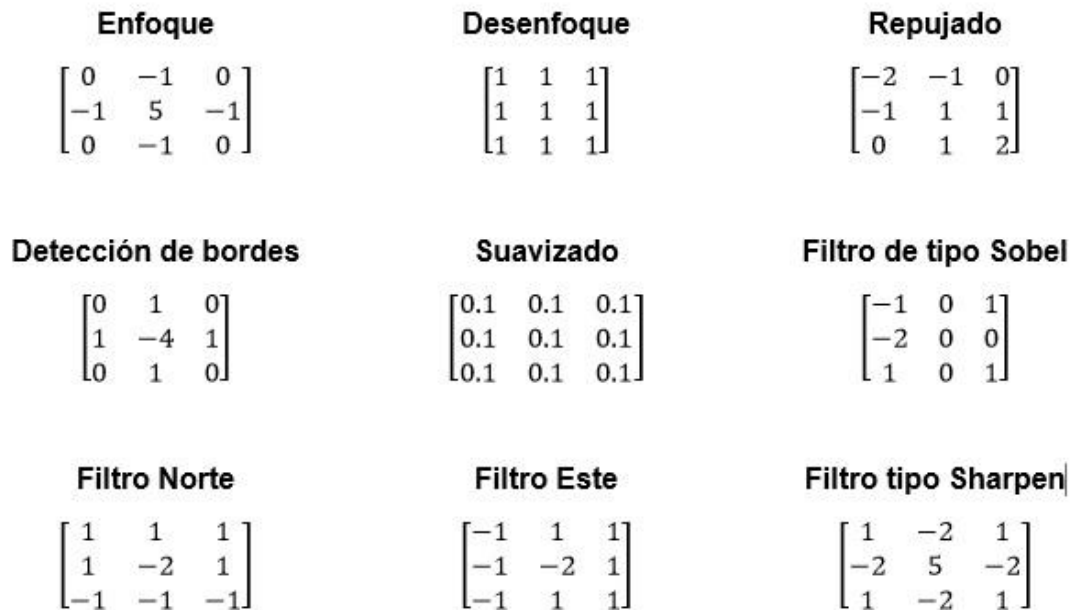


Figura 4. Máscaras más usadas.

En la referencia (PHP, 2005) se encuentra un sencillo simulador del efecto que tiene sobre una imagen fija una matriz de convolución dada (Giménez, Serrá, Martínez, 2016).

Otra de las técnicas muy conocidas que forma parte del procesamiento digital de imágenes en la búsqueda de mejorar sus propiedades y facilitar la extracción de características es la conversión a escala de grises de imágenes a color para posterior binarización. Dentro de esta práctica se encuentran tres maneras de convertir una imagen a color en escala de grises, la primera es la forma más simple y se trata de sumar las componentes RGB de cada capa pixel a pixel y dividir entre la cantidad de total de unidades (tres), es decir un promedio.

$$M_Grey_{(i,j)} = \frac{(R_{(i,j)}+G_{(i,j)}+B_{(i,j)})}{3} \quad (2)$$

Como se mencionó antes, la imagen es tratada como un arreglo bidimensional donde x e y representan las coordenadas de cada pixel, para la **Ecuación 2** $M_Grey_{(i,j)}$ es la matriz resultante de realizar el promedio de los componentes RGB de la imagen procesada. La segunda forma es la tonalidad, que en el modelo HSL (del inglés Hue, Saturation, Lightness con su equivalencia en español Tonalidad, Saturación, Luminancia o intensidad) se define como parte de la media de las componentes de color máxima y mínima. Esta definición coloca los colores primarios y secundarios en un segundo plano, el método afecta de manera importante el contraste de la imagen y se representa mediante la **Ecuación 3**.

$$M_Grey_{(i,j)} = \frac{(Max[R_{(i,j)},B_{(i,j)}])+(Min[R_{(i,j)},G_{(i,j)},B_{(i,j)}])}{2} \quad (3)$$

La tercera manera es la Luminosidad, ésta es la versión sofisticada del promedio, los valores de cada color presentan un valor teniendo en cuenta la percepción humana siendo definida por la **Ecuación 4** (Jimmy & Cortés, 2011)

$$M_Grey_{(i,j)} = 0,299 * R_{(i,j)} + 0.587 * G_{(i,j)} + 0.114 * B_{(i,j)} \quad (4)$$

El escalado de imágenes también pertenece al grupo de procesamiento digital, éste modifica la relación espacial entre píxeles y consiste en dos operaciones básicas, la primera define la reubicación de los píxeles en el plano imagen y la segunda es una interpolación de los niveles de grises, los cuales tiene que ver con la asignación de los valores de intensidad de los píxeles en la imagen transformada. El método de interpolación de niveles de grises más simple es el fundamentado en el vecino más cercano, que consiste en escoger como el valor del pixel de una imagen de salida, el valor del pixel de entrada más cercano en la dirección inversa. Este método, aunque es simple de interpretar, presenta frecuentemente características indeseables en las imágenes, tales como distorsión de líneas en imágenes de alta resolución. Este tipo de interpolación logra producir un error espacial relativo a no más de $\frac{1}{\sqrt{2}}$ pixel por unidad. En la **Figura 5** se ilustra el método del vecino más cercano en la interpolación de niveles de grises. El esquema consiste en tomar el pixel de la posición $(x, y) = (a, b)$ en la imagen original y transformarlo mediante las **Ecuaciones 5 y 6** de distorsión. (Melo, n.d.)

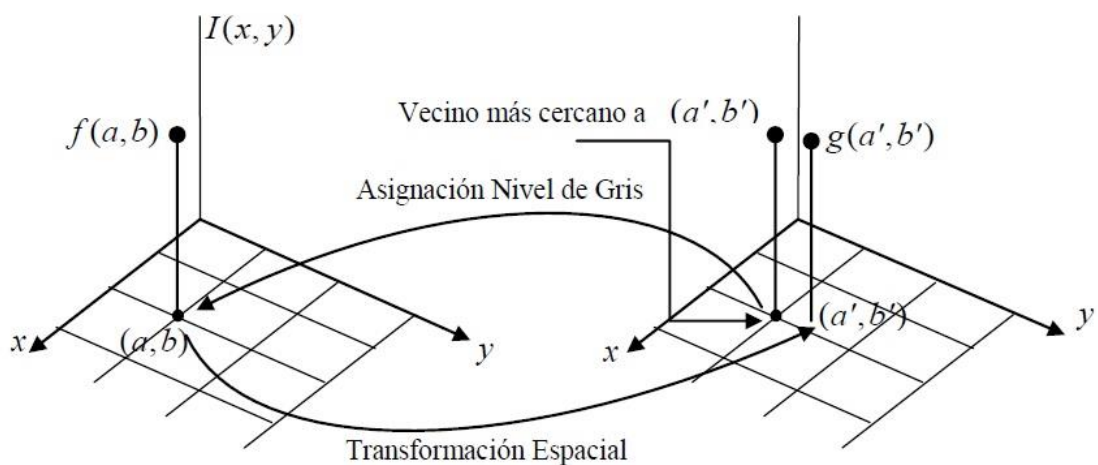


Figura 5. Interpolación de nivel de gris basado en el concepto del vecino más cercano.

$$x' = r(x, y) = c_1x + c_2y + c_3xy + c_4 \quad (5)$$

$$y' = s(x, y) = c_5x + c_6y + c_7xy + c_8 \quad (6)$$

La binarización o umbralización de imágenes es uno de los métodos que abarca la segmentación de imágenes digitales. Las técnicas de umbralización buscan obtener un valor de umbral que permita binarizar la imagen separando adecuadamente el fondo (*Background*) y el objeto a separar (*Foreground*). Muchas de las técnicas de umbralización están basadas en la información estadística que brinda el histograma, sobre todo en aquellas imágenes donde los objetos tienen una superficie o textura homogénea y el fondo es más o menos uniforme. El problema de la binarización es encontrar el valor T (umbral) adecuado entre los valores de grises en imágenes que permita una óptima separación entre el objeto y el fondo. En un caso ideal, el histograma de una imagen posee dos picos o lóbulos, se podrían separar dos zonas o regiones y el umbral será aquel valor que se encuentra en el valle entre ambas. Sin embargo, las imágenes siempre presentan factores como el ruido, la luz, la oclusión, el traslape, entre otros y como consecuencia, sus histogramas son difíciles de analizar, es decir, presentan histogramas sin una marcada separación de regiones. Existen numerosas técnicas de umbralización desarrolladas, ampliadas y modificadas por diversos autores pero ninguna representa un método general que se aplique a toda las imágenes y que realice una binarización perfecta, dentro de la gama de estos métodos, los más representativos son aquellos basados en la forma del histograma, en la clusterización, en la entropía, en los atributos de la imagen, en información espacial y en características locales. La umbralización basada en la entropía ha recibido una consideración importante cuyo principio es medir la incertidumbre que describe la información contenida en una fuente, su pionero fue Pun (1980, 1981) quien considero los pixeles de una imagen convertida a 256 niveles de gris y se separan en dos niveles principales de gris, el primer plano o (*Foreground*) y un fondo de base o (*Background*). La variable g denotará estos

valores de niveles de gris. Para imágenes de 8 bits $g = 0 \dots 255$. Sea el conjunto de pixeles de la imagen:

$$I = \{g \in I / 0 \leq g \leq 255\} \quad (7)$$

Pixeles correspondientes al objeto (*Foreground*) y el fondo (*Background*):

$$F = \{g \in I / g > t\} \quad (8)$$

$$B = \{g \in I / g \leq t\} \quad (9)$$

En el contexto de procesamiento de imágenes, el objeto es el conjunto de pixeles con luminancia menor a un cierto valor T , mientras que el fondo base es el conjunto de pixeles cuya luminancia supera este valor (Cattaneo, Larcher, Ruggeri, Herrera, BIASONI, 2011).

Para completar un sistema inteligente, solo falta definir los métodos de inteligencia artificial aplicables al sistema desarrollado.

Como primer método, están las redes neuronales artificiales (RNA) que pueden definirse en numerosas formas desde definiciones cortas y genéricas hasta aquellas que intentan explicar detalladamente su significado, (Kohonen, 88c) las define como “redes interconectadas masivamente en paralelo de elementos simples (usualmente adaptativos) y con organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema nervioso biológico”. Es necesario destacar que tales computadores neuronales no ejecutan las típicas instrucciones de máquina de los computadores digitales, a menos que estén hechos para emular el comportamiento de las redes neuronales físicas. Debido a su constitución y fundamentos, las RNA presentan un gran número de características semejantes a las del cerebro. Por ejemplo, son capaces de aprender de la experiencia, de generalizar de casos anteriores a nuevos casos, de abstraer características esenciales a partir de entradas que representan información

irrelevante, etc. Esto hace resaltar numerosas ventajas como el aprendizaje adaptativo con la capacidad de aprender a realizar tareas basadas en un entrenamiento o una experiencia inicial, la autoorganización donde una red puede crear su propia representación de la información que recibe mediante un etapa de aprendizaje, la tolerancia a fallos que si existe la destrucción parcial de una red y conduce a la degradación de su estructura, algunas capacidades se pueden retener, la operación en tiempo real ya que los computadores neuronales pueden ser realizados en paralelo y la fácil inserción dentro de la tecnología existente. Las RNA son una tecnología computacional emergente que puede utilizarse en un gran número y variedad de aplicaciones, tanto comerciales como militares, algunas áreas donde existen aplicaciones de las RNA son la biología, el medio ambiente, las finanzas, la manufacturación, la medicina y los militares.

Dentro de las RNA se destaca la red Backpropagation ya que al tener un algoritmo de propagación hacia atrás o retro propagación, es una regla de aprendizaje que se puede aplicar en modelos de redes con más de dos capas de células. Una característica importante de este algoritmo es la representación interna del conocimiento que es capaz de organizar en la capa intermedia de las células para conseguir cualquier correspondencia entre la entrada y la salida de la red. El funcionamiento de una RNA Backpropagation consiste en un aprendizaje de un conjunto predefinido de pares entradas-salidas dados como ejemplo, empleando un ciclo propagación – adaptación de dos fases: primero se aplica un patrón de entrada como estímulo para la primera capa de las neuronas de la red, se va propagando a través de todas las capas superiores hasta generar una salida, se compara el resultado en las neuronas de salida con las neuronas que se desean obtener y se calcula el valor del error para cada neurona de salida. Seguidamente, estos errores se transmiten hacia atrás, partiendo de la capa de salida hacia todas las neuronas de la capa intermedia que contribuyan directamente a la salida, recibiendo el porcentaje de error aproximado a la participación de la neurona intermedia en la salida original. Este proceso se repite capa por capa hasta que todas las neuronas de la red hayan recibido un error que describa su aportación relativa al error total.

Basándose en el valor del error recibido, se reajustan los pesos de conexión de cada neurona, de manera que en la siguiente vez que se presente el mismo patrón, la salida esté más cercana a la deseada, es decir, que el error disminuya.

Para aplicar el algoritmo de entrenamiento *Backpropagation* es recomendable seguir estos pasos:

1. Inicializar los pesos de la red con valores pequeños aleatorios.
2. Presentar un patrón de entrada $X_p: X_{p1}, X_{p2}, \dots, X_{pN}$, y especificar la salida deseada que debe generar la red d_1, d_2, \dots, d_M . Si la red se utiliza como clasificador, todas las salidas deseadas serán cero, salvo una, que será la de la clase a la cual pertenece).
3. Calcular la salida actual de la red, para ello se presentan las entradas la red y se va calculando la salida de la red y_1, y_2, \dots, y_M los pasos son los siguientes:

- a. Se calculan las entradas netas para las neuronas ocultas procedentes de las neuronas de entrada.

Para una neurona j oculta:

$$net_{pj}^h = \sum_{i=1}^N w_{ji}^h x_{pi} + \theta_j^h \quad (10)$$

En donde el índice h se refiere a magnitudes de la capa oculta (hidden); el subíndice p , al p -ésimo vector de entrenamiento y j a la j -ésima neurona oculta. El Término θ puede ser opcional, pues actúa como una entrada más.

- b. Se calculan las salidas de las neuronas ocultas:

$$y_{pj} = f_j^h(net_{pj}^h) \quad (11)$$

- c. Se realizan los mismos cálculos para obtener las salidas de las neuronas de salida

$$net_{pk}^0 = \sum_{j=1}^L w_{kj}^0 y_{pj} + \theta_k^0 \quad (12)$$

$$y_{pk} = f_k^0(net_{pk}^0) \quad (13)$$

4. Calcular los términos de error para todas las neuronas. Si la neurona k es una neurona de la capa de salida, el valor de la delta es:

$$\delta_{pk}^o = (d_{pk} - y_{pk})f_k^o(\text{net}_{pk}^o) \quad (14)$$

5. Actualización de pesos, para ello se utiliza el algoritmo recursivo, comenzando por las neuronas de salida y trabajando hacia atrás hasta llegar a la capa de entrada, ajustando los pesos así:

Para los pesos de las neuronas de la capa de salida:

$$w_{kj}^o(t + 1) = w_{kj}^o + \Delta w_{kj}^o(t + 1); \quad (15)$$

$$\Delta w_{kj}^o(t + 1) = \alpha \delta_{pk}^o y_{pj} \quad (16)$$

Y para los pesos de las neuronas de la capa oculta

$$w_{ji}^h(t + 1) = w_{ji}^h + \Delta w_{ji}^h(t + 1); \quad (17)$$

$$\Delta w_{ji}^h(t + 1) = \alpha \delta_{pj}^o x_{pi} \quad (18)$$

6. El proceso se repite hasta que el término de error

$$E_p = \frac{1}{2} \sum_{k=1}^M \delta_{pk}^2 \quad (19)$$

Resulte aceptablemente pequeño para cada uno de los patrones aprendidos (Hilera, Martinez, 1994).

Ahora bien, un Modelo Oculto de Markov HMM es un modelo de secuencia o clasificador de secuencia cuyo trabajo es asignar etiquetas de clases a cada unidad en una secuencia, mapeando así una secuencia de observaciones. Los HMM son modelos de serie temporal que asumen dependencia entre el estado actual de los datos de series de tiempo y su estado anterior. Los estados de los datos están ocultos y se pueden estimar a partir de la asociación probabilística entre los datos

observados y los estados, así como de las probabilidades de transición entre los estados en el tiempo (Chuk, Crookes, Hayward, Chan, & Hsiao, 2017).

Como un HMM es un modelo probabilístico donde dada una secuencia de unidades (palabras, oraciones, letras) calculan una distribución de probabilidad sobre posibles secuencias elige la mejor. Uno de los algoritmos de entrenamiento para los HMM es el *Forward-Backward* o Avance-Retroceso que se basa en aprender los parámetros HMM A y B, dada una secuencia de observación O y el conjunto de posibles estados en HMM. La entrada a este algoritmo sería una secuencia no etiquetada de observaciones y un vocabulario de posible estados ocultos P. Este algoritmo permitirá entrenar tanto las probabilidades de transición A como las probabilidades de emisión B, crucialmente al ser iterativo funciona calculando una estimación inicial para luego usarlas y calcular una mejor estimación y así sucesivamente mejorando con cada ciclo las probabilidades de aprendizaje.

Para comprender el algoritmo, es necesario definir una probabilidad útil relacionada con la probabilidad hacia atrás β que es la probabilidad de ver las observaciones de tiempo $(t + 1)$ hasta el final, dado que se encuentra en estado i en el tiempo t :

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, O_T | qt = i, \lambda) \quad (20)$$

El algoritmo *Forward-Backward* comienza con una estimación inicial de los parámetros HMM $\lambda = (A, B)$. Luego se ejecutan dos pasos, iterativamente, el paso de expectativa o *E-step* y el paso de maximización o *M-step*. En el primer paso se calcula e recuento de ocupación de estado esperado γ y el estado de transición ξ de las probabilidades A y B anteriores; y en el segundo paso se usan estos valores para recalcular las nuevas probabilidades A y B,

A continuación se presenta el algoritmo *Forward-Backward* planteado por (Jurafsky & Martin, 2017) el cual recibe unas observaciones de tamaño T , un vocabulario de salida V y un conjunto de estados ocultos Q para retornar los estados de transición y probabilidad.

1. Inicializar A y B con valores aleatorios entre cero (0) y uno (1).
2. Iniciar la iteración hasta la convergencia ya sea por la cantidad de observaciones o un valor límite establecido.

3. E-Step

$$\gamma_t(j) = \frac{\alpha_t(j)\beta_t(j)}{\alpha_T(q_F)} \quad \forall t \text{ and } j \quad (21)$$

$$\xi_t(i, j) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\alpha_T(q_F)} \quad \forall t, i \text{ and } j \quad (22)$$

4. M-Step

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \sum_{k=1}^N \xi_t(i, k)} \quad (23)$$

$$b_{i(v_k)} = \frac{\sum_{t=1}^T \text{s.t. } o_t=v_k \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad (24)$$

5. Retorno de A y B

Para el almacenamiento de los datos se utilizó MySQL® como motor de base de datos, este es un sistema de gestión de bases de datos relacional en el cual cualquier persona puede contribuir para incluir elementos, arreglar problemas, realizar mejoras o sugerir optimizaciones. Las bases de datos por lo general son compartidas y se emplean para almacenar datos persistentes, dichos datos pueden considerarse como una representación de las entidades, junto con los vínculos que están entre estas. Uno de los beneficios más importantes que ofrecen los sistemas de bases de datos es el de la independencia de los datos, esta requiere que se haga una clara distinción entre el modelo de datos y su implementación. Los sistemas de bases de datos también soportan por lo regular transacciones o unidades de trabajo lógicas, una ventaja de las transacciones es que se garantiza que sean atómicas es decir se ejecuta todo o nada, incluso si el sistema falla a la mitad de su ejecución. Los sistemas de bases de datos pueden estar fundamentados en varias teorías

diferentes, en particular los sistemas relacionales se basan en una teoría formal denominada modelo relacional, según la cual los datos están relacionados como filas de tablas, y cuentan con operadores que manejan directamente el proceso de inferir proposiciones verdaderas adicionales a partir de las ya dadas (Acosta, Nieto, Barahona, 2015).

1.7.3. MARCO LEGAL

En el año 1991 nace la norma ISO 9126 la cual recibe como nombre “Information technology-Software Product Evaluation-Quality characteristics and guidelines for their use” (Información de productos de tecnología de software características evaluación de calidad y directrices para su uso), dicha norma está dividida en cuatro partes las cuales son: realidad, métricas externas, métricas internas y calidad en las 26 métricas de uso. En donde la primera parte de esta estandarización divide la calidad en distintas características y subcaracterísticas de la siguiente manera:

- Funcionalidad: adecuación, exactitud, interoperabilidad, seguridad y cumplimiento funcional.
- Fiabilidad: Madurez, recuperabilidad, tolerancia a fallos y cumplimiento de fiabilidad.
- Usabilidad: Aprendizaje, comprensión y operatividad.
- Eficiencia: comportamiento en el tiempo y comportamiento de los recursos.
- Mantenibilidad: Estabilidad, facilidad de análisis, facilidad de cambio y facilidad de pruebas.
- Portabilidad: capacidad de instalación, capacidad de reemplazamiento, adaptabilidad y coexistencia.

Este conjunto de normas (ISO 9126 Y 14598) inicia su constitución en 1998 y concluye en el 2004.

Para este entonces ya ha surgido en el 2000 la reglamentación ISO 25000 “conocida como SQuaRE (System and Software Quality Requirements and Evaluation), es una

familia de normas que tiene por objetivo la creación de un marco de trabajo común para evaluar la calidad del producto software” (Portal oficial ISO 25000

II. DOCUMENTACIÓN DEL SOFTWARE

2.1. PLAN DE PROYECTO

Este proyecto fue ideado inicialmente para desarrollarlo y postularlo como opción de grado únicamente, hoy día y después de tantos logros con este tema, se ha pensado en continuar con esta gran investigación y seguir el desarrollo como proyecto personal y de presentarse opciones de trabajo gracias a este proyecto, bienvenidas sean.

2.2. DETERMINACION DE REQUERIMIENTOS

| | |
|--|---|
| Identificación del requerimiento: | RF01 |
| Nombre del Requerimiento: | Levantamiento de Información. |
| Características: | Recopilación de información sobre las mariposas. |
| Descripción del requerimiento: | Se recibirá por parte del grupo CIAM el archivo digital con la información pertinente y detallada de cada una de las siete especies de mariposas que actualmente habitan el parque Metropolitano María Lucia. |
| Prioridad del requerimiento: | Alta |

Tabla 2. *Requerimiento Funcional 1*

| | |
|--|------|
| Identificación del requerimiento: | RF02 |
|--|------|

| | |
|---|--|
| Nombre del Requerimiento: | Diseño de la Herramienta |
| Características: | Diseño del modelado para el PSI |
| Descripción del requerimiento: | Se hará un diseño de interfaz amigable, sencilla de manejar e intuitiva. |
| Prioridad del requerimiento: Alta | |

Tabla 3. *Requerimiento Funcional 2*

| | |
|---|--|
| Identificación del requerimiento: | RF03 |
| Nombre del Requerimiento: | Inicio de la Arquitectura del Proyecto |
| Características: | Creación del login. |
| Descripción del requerimiento: | Este primer paso determinará qué tipo de usuario ingresa al sistema (Administrador/Usuario general). |
| Prioridad del requerimiento: Alta | |

Tabla 4. *Requerimiento Funcional 3*

| | |
|--|-----------------------------------|
| Identificación del requerimiento: | RF04 |
| Nombre del Requerimiento: | Módulo de Administrador. |
| Características: | Especificación Rol Administrador. |

| | |
|---------------------------------------|--|
| Descripción del requerimiento: | <p>El administrador es la persona encargada de gestionar el sistema en cuanto al ingreso de información se refiere y tendrán las siguientes funciones:</p> <ul style="list-style-type: none"> • Ingresar, modificar, eliminar y consultar información de las mariposas según sea necesario. • Ejecutar el entrenamiento del PSI ya sea de la RNA y del HMM. • Ejecutar la aplicación del sistema de reconocimiento. |
| Prioridad del requerimiento: | |
| Alta | |

Tabla 5. Requerimiento Funcional 4

| | |
|--|---|
| Identificación del requerimiento: | RF05 |
| Nombre del Requerimiento: | Modulo Usuario General |
| Características: | Especificación Rol Usuario. |
| Descripción del requerimiento: | <p>Este rol tendrá la siguiente función:</p> <ul style="list-style-type: none"> • Ejecutar el sistema inteligente identificando mariposas. |
| Prioridad del requerimiento: | |
| Alta | |

Tabla 6. Requerimiento Funcional 5

| | |
|--|--------------------------------|
| Identificación del requerimiento: | RF06 |
| Nombre del Requerimiento: | Presentación de la Información |

| | |
|---|--|
| Características: | Resultados del reconocimiento |
| Descripción del requerimiento: | Al momento que el sistema reconozca una mariposa, la información taxonómica y ecológica básica de la mariposa se visualizará en forma gráfica y textual. |
| Prioridad del requerimiento: Alta | |

Tabla 7. Requerimiento Funcional 6

| | |
|---|--|
| Identificación del requerimiento: | RF07 |
| Nombre del Requerimiento: | Modulo Almacenamiento de la Información. |
| Características: | Crear la Base de Datos. |
| Descripción del requerimiento: | Desarrollar una base de datos local con la información de las mariposas presentada en el documento digital dado por el CIAM. |
| Prioridad del requerimiento: Alta | |

Tabla 8. Requerimiento Funcional 7

| | |
|--|---|
| Identificación del requerimiento: | RF08 |
| Nombre del Requerimiento: | Módulo – Reconocimiento de Patrones #1. |
| Características: | Red Neural Artificial (RNA) |

| | |
|---|--|
| Descripción del requerimiento: | Diseñar la estructura de la RNA e implementarla en el PSI. |
| Prioridad del requerimiento: Alta | |

Tabla 9. Requerimiento Funcional 8

| | |
|---|--|
| Identificación del requerimiento: | RF09 |
| Nombre del Requerimiento: | Módulo – Reconocimiento de Patrones #2. |
| Características: | Modelos Ocultos de Markov (HMM) |
| Descripción del requerimiento: | Diseñar la estructura del HMM e implementarlo en el PSI. |
| Prioridad del requerimiento: Alta | |

Tabla 10. Requerimiento Funcional 9

| | |
|--|---|
| Identificación del requerimiento: | RF010 |
| Nombre del Requerimiento: | Módulo de Comparación. |
| Características: | Enfrentamiento de los métodos de reconocimiento. |
| Descripción del requerimiento: | En este módulo se comparan los resultados obtenidos de cada método de reconocimiento de patrones y se determinará cuál de ellos responde mejor en la identificación de las mariposas. |
| Prioridad del requerimiento: | |

Tabla 11. Requerimiento Funcional 10

2.3. ESPECIFICACIÓN DEL SISTEMA

- **Políticas de la Empresa:** Todo el material que se realice para el usuario y el sistema como tal deben estar en español. Las fotografías que se carguen en el sistema deberán tener un mínimo de 1024x768 pixeles y en formato JPEG. El sistema deberá tener un diseño e implementación sencilla, independiente de la plataforma o del lenguaje de programación.
- **Limitaciones de Hardware:** Para este software será necesario un computador en el cual se instalará MySQL® y JAVA®.
- **Interfaces con Otras Aplicaciones:** Debido a que REIDMAR no interactúa con otros sistemas y es autónomo no se desarrollaran interfaces con otras aplicaciones.
- **Lenguajes de Programación:** Sistema de Gestión de Base de Datos MySQL® y lenguaje de Programación JAVA® en el entorno de desarrollo Netbeans IDE 7.4®.
- **Protocolos de Comunicación:** Las conexiones necesarias para la utilización de MySQL® y JAVA®, se harán por medio de la configuración de estos programas, por lo tanto no es necesario implementar algún tipo de protocolo de comunicación en especial.
- **Criticidad de la Aplicación:** Para garantizar la buena credibilidad, el sistema deberá ser sometido a una serie de pruebas para establecer que se encuentra acorde a los requerimientos que se plasman en el documento en tanto a la consistencia de datos como al rendimiento del PSI.
- **Consideraciones Acerca de la Seguridad:** Para el acceso al PSI del administrador, la clave de seguridad deberá estar segura y en su defecto

encriptada en la base de datos para brindar buena seguridad al sistema y su información.

2.4. ESPECIFICACIÓN DEL DISEÑO

2.4.1. MODELO DE ENTIDAD RELACIÓN (MER)

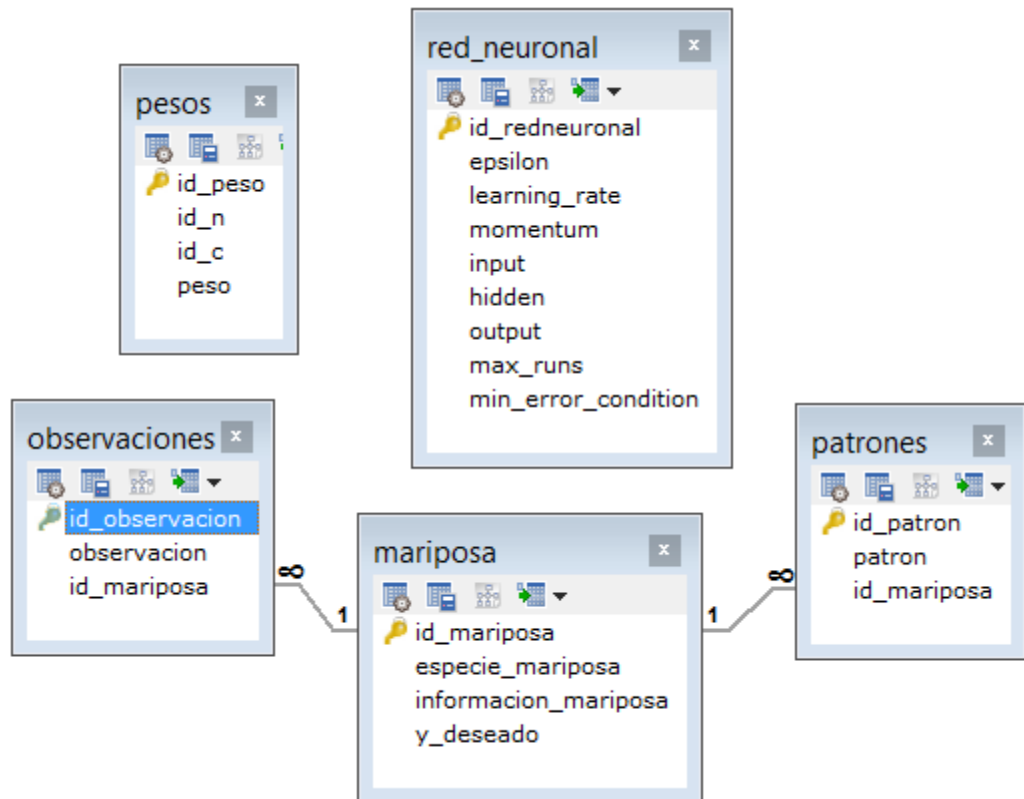


Figura 6. Modelo Entidad Relación Base de Datos

2.4.2. DIAGRAMAS DE CASOS DE USO

Los diagramas de casos de uso “modelan la funcionalidad del sistema según lo perciben los usuarios externos. Cada caso de uso corresponde a una unidad coherente de funcionalidad, expresada como transición entre los actores y el sistema. El propósito de los diagramas de casos de uso es enumerar los actores y

casos de uso, demostrando que actor(es) participa(n) en cada caso de uso.” (Rumbaugh, Jacobson, Booch, 2000).

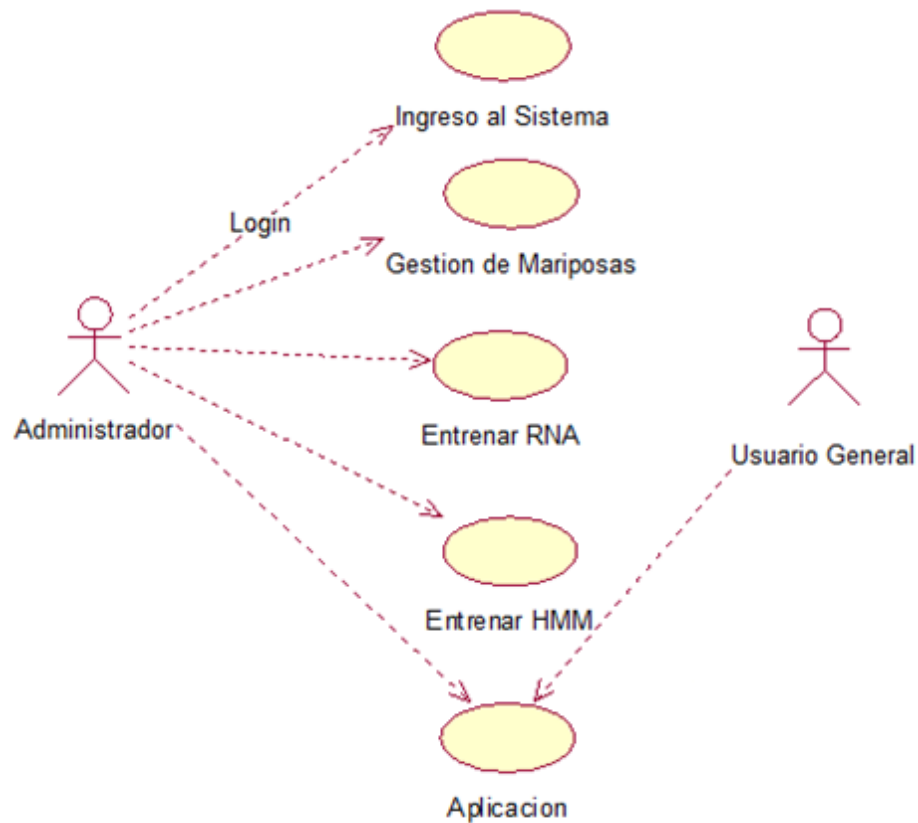


Figura 7. Casos de Uso

DESCRIPCIÓN CASOS DE USO

Rol Administrador

- Ingreso al Sistema: Mediante un *login* el usuario administrador accede al sistema, de lo contrario será tratado como un usuario general.
- Gestión de Mariposas: El administrador tendrá la potestad de registrar, editar y eliminar mariposa(s).
- Entrenar RNA y HMM: En este caso de uso, el administrador podrá dar la orden a la herramienta de iniciar el entrenamiento de la Red Neuronal

Artificial con los patrones guardados y el Modelo Oculto de Markov con las observaciones almacenadas en el sistema.

- Entrenar HMM: Al igual que el ítem anterior pero en este se cargar las observaciones a los Modelos Ocultos de Markov.
- Aplicación: El administrador comprobara si los entrenamientos fueron exitosos al intentar identificar nuevas fotografías de las mariposas.

Usuario General

- Aplicación: Verificación del aprendizaje del sistema inteligente de parte de cualquier usuario donde validan el correcto aprendizaje de la RNA y del HMM.

2.4.3. DIAGRAMAS DE SECUENCIAS

Según (Rumbaugh, Jacobson, Booch, 2000) “Un diagrama de secuencia muestra un conjunto de mensajes, dispuestos en una secuencia temporal. Cada rol en la secuencia se muestra como una línea de vida, es decir, una línea vertical.

Ingreso al Sistema

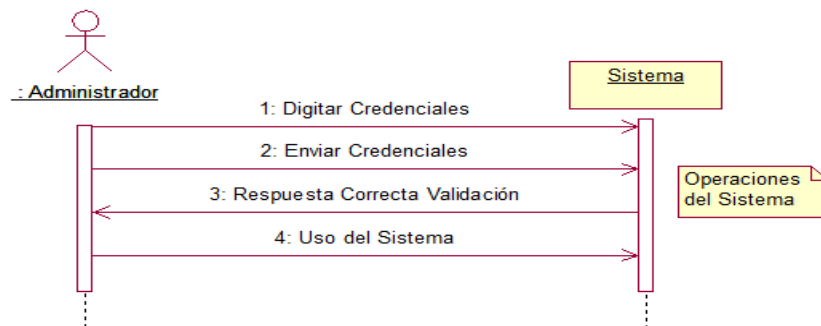


Figura 8. Diagrama de Secuencia Ingreso al Sistema

El Administrador abre el panel del *Login* y una vez allí digita las credenciales de acceso (Usuario/Contraseña), luego la envía y espera respuesta de la validación del sistema, accediendo a las configuraciones de administrador si las credenciales son correctas de lo contrario no podrá ingresar.

Gestión de Mariposas

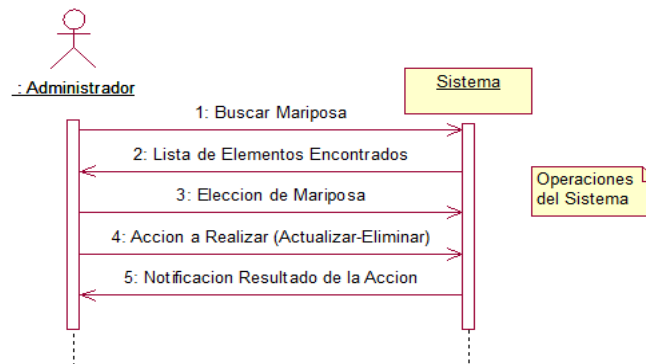


Figura 9. Diagrama de Secuencia Gestión Mariposa

El administrador tiene la posibilidad de buscar mariposa, actualizar datos, borrar y registrar mariposa.

Entrenamiento RNA y HMM

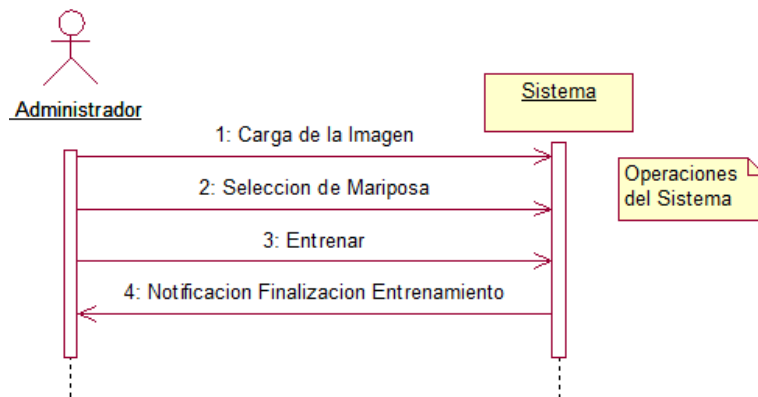


Figura 10. Diagrama de Secuencia Entrenamiento Sistema

Para entrenar el sistema, el administrador solo debe cargar la imagen, seleccionar la mariposa correspondiente, iniciar el entrenamiento y esperar que el sistema notifique la terminación del proceso.

Aplicación Administrador

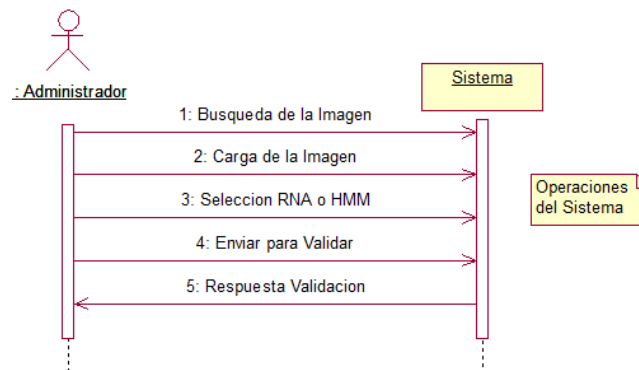


Figura 11. Diagrama de Secuencia Uso del Sistema Administrador

Inmediatamente después del entrenamiento el administrador puede hacer el proceso de validación y aceptación de las respuestas del sistema.

Aplicación Usuario General

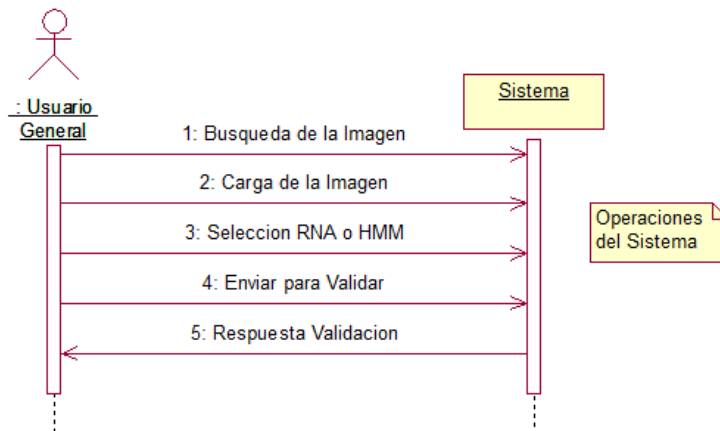


Figura 12. Diagrama de Secuencia Aplicación Usuario

El Usuario general puede validar el sistema seleccionando la imagen de la mariposa a evaluar mediante las dos opciones, por RNA o por HMM.

2.4.4. DIAGRAMAS DE ACTIVIDADES

El diagrama de actividades muestra las operaciones computacionales implicadas en la ejecución de un cálculo. (Rumbaugh, Jacobson, Booch, 2000).

Ingreso al Sistema

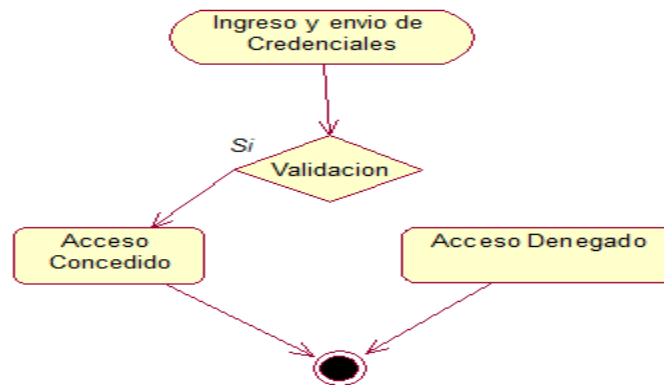


Figura 13. Diagrama de Actividades Ingreso.

Gestión de Mariposas

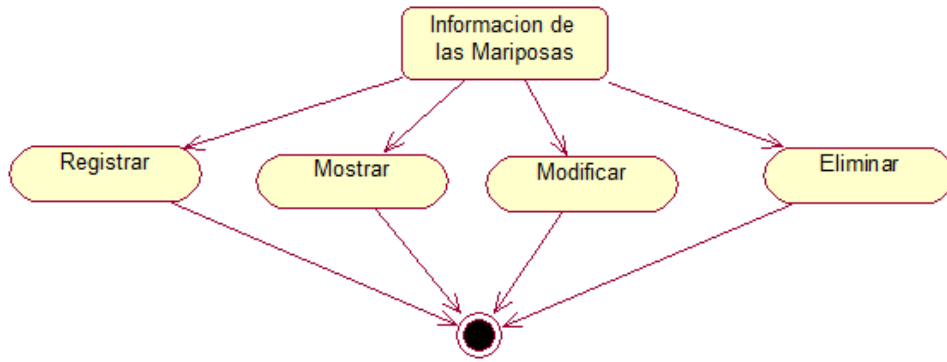


Figura 14. Diagrama de Actividades Mariposas.

Entrenamiento RNA y HMM

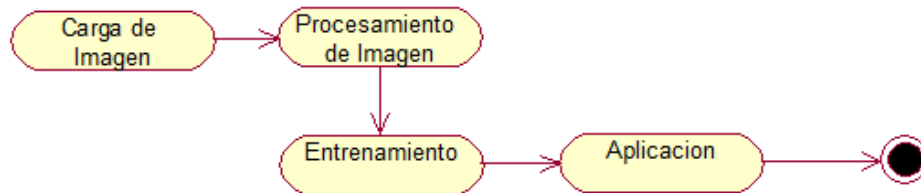


Figura 15. Diagrama de Actividades Entrenamiento.

Aplicación Administrador y Usuario General

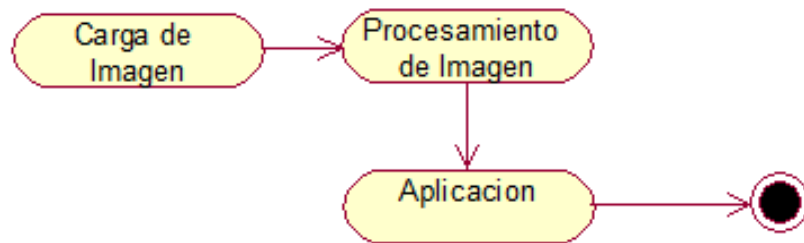


Figura 16. Diagrama de Actividades Usuarios.

2.4.5. DIAGRAMA DE CLASES

Para (Rumbaugh, Jacobson, Booch, 2000) un diagrama de clases representa el centro alrededor del cual se organiza el sistema en general,

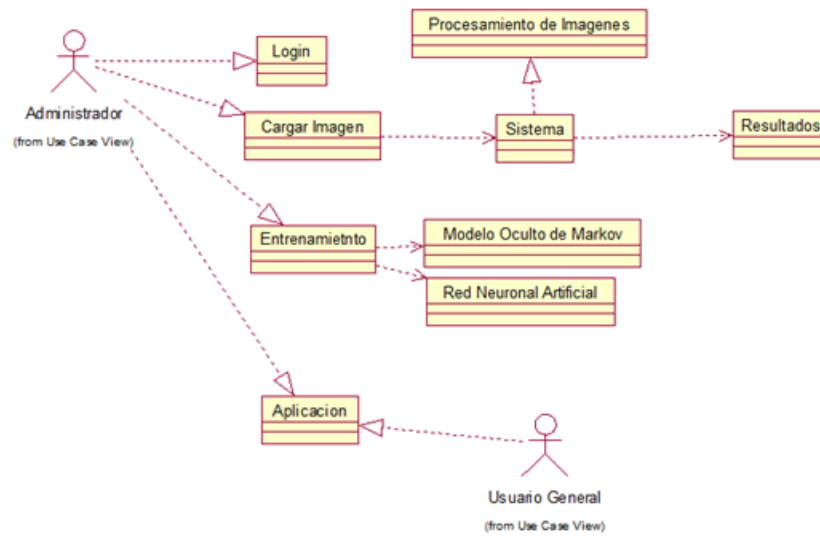


Figura 17. Diagrama de Clases.

En este diagrama de clases se encuentran las principales clases utilizadas en el software y su ciclo/flujo para cada clase. La clase llamada sistema hace referencia a la parte que el usuario ya sea administrador o usuario general no controla de la aplicación en ejecución.

2.4.6. DIAGRAMA DE DESPLIEGUE

Según (Rumbaugh, Jacobson, Booch, 2000), un diagrama de despliegue “Representa la disposición de las instancias de componentes de ejecución en instancias de nodos o recursos de ejecución, al igual que una computadora, dispositivo o memoria y permite determinar el impacto de la distribución y asignación de recursos.”

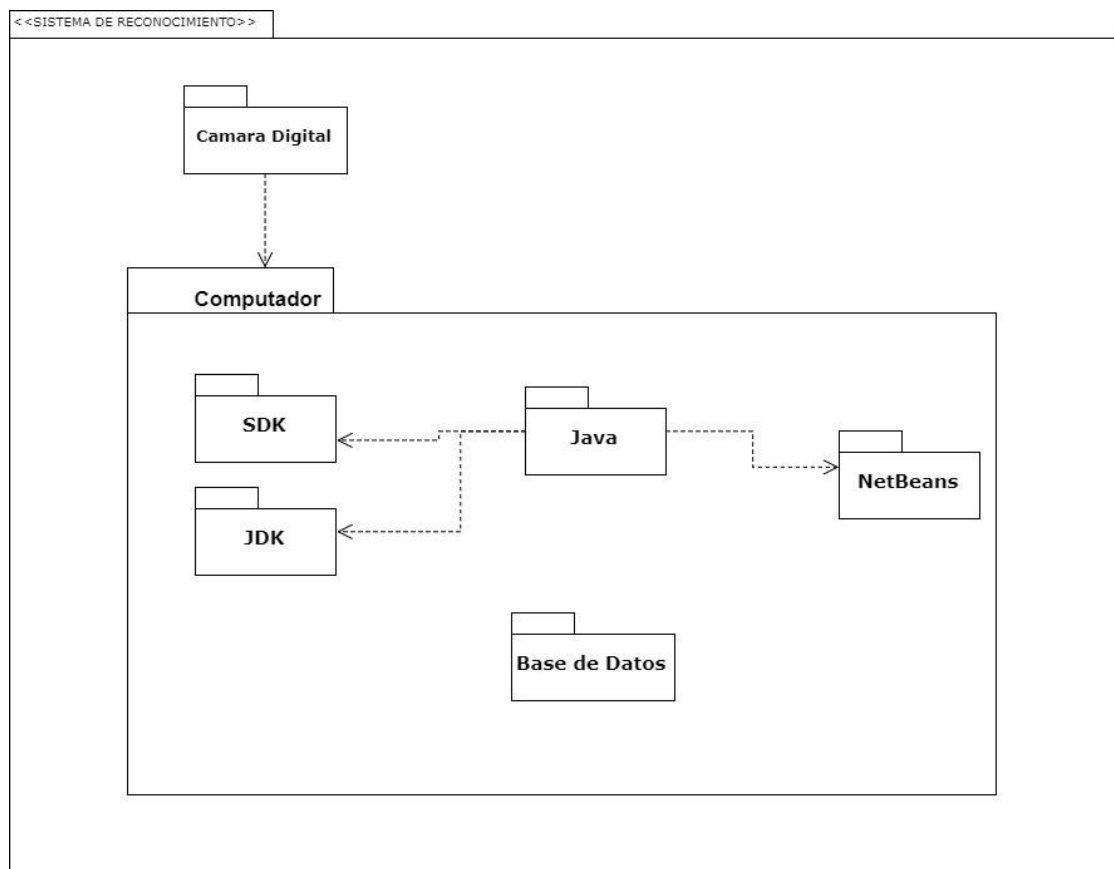


Figura 18. Diagrama de Despliegue.

III. DOCUMENTACION DEL SISTEMA INTELIGENTE

3.1. PROCESAMIENTO DE IMÁGENES

Como se explicó en la metodología, la **Figura 19** representa las fases del sistema desarrollado. En este módulo se explicarán con detalle cada una de ellas y se mostrarán paralelamente capturas de pantalla de la herramienta llamada Registro e Identificación de Mariposas o por sus siglas REIDMAR v1.0.

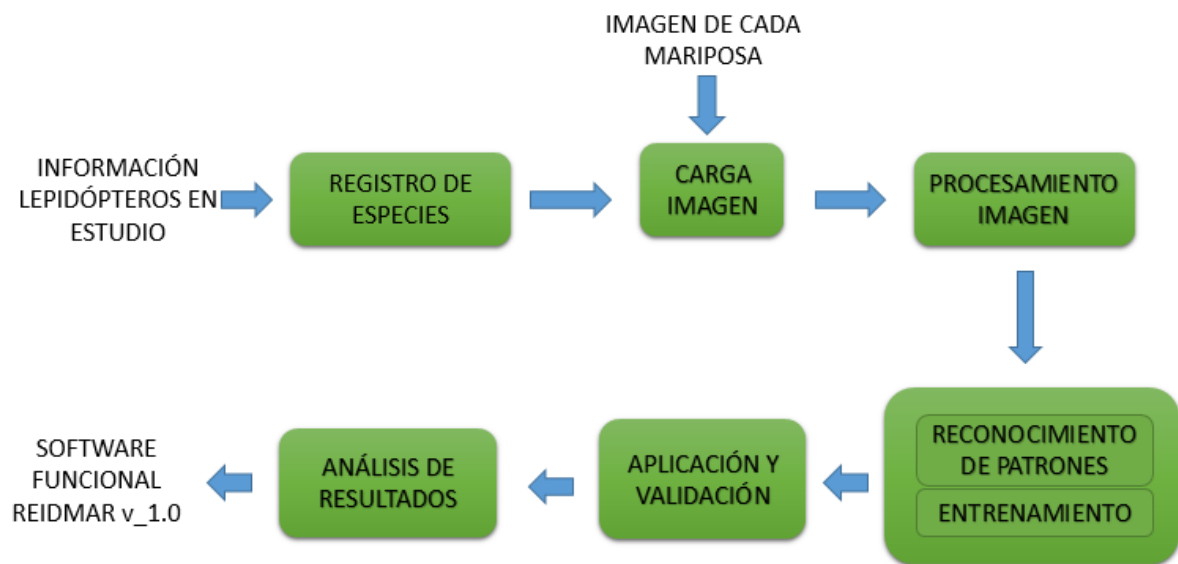


Figura 19. Metodología del proyecto.

En la primera fase se registraron las cinco especies de mariposas a identificar, utilizando el módulo de gestión de mariposas del aplicativo REIDMAR **Figura 25**.

Las mariposas ingresadas al sistema y algunas de sus características fueron las siguientes:

✓ *Methona Confusa*

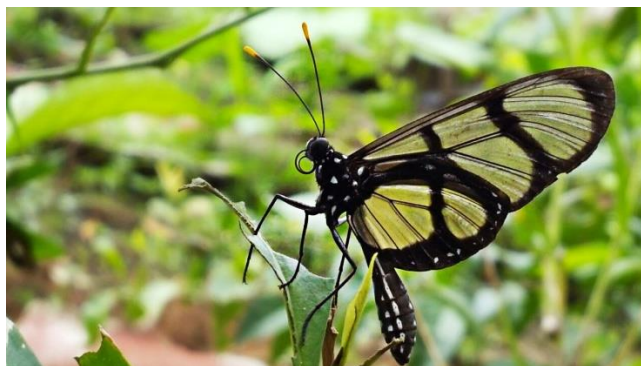


Figura 20. Foto de mariposa *Methona Confusa* en el PMML.

Orden: Lepidoptera.

Familia: Nymphalidae.

Género: Methona.

Especie: Methona Confusa.

Planta Hospedera: Brunfelsia australis - Francesina.

✓ *Dione Juno*

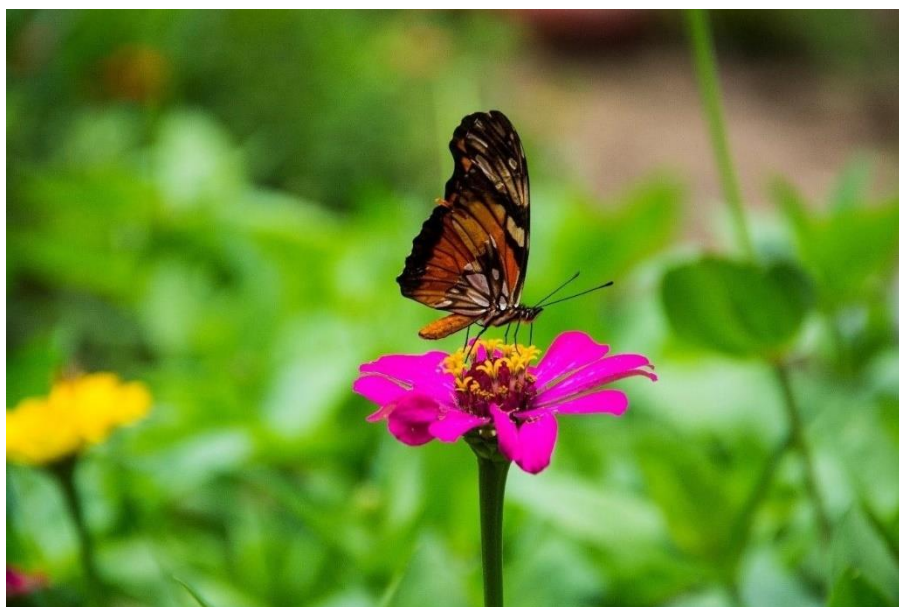


Figura 21. Foto de mariposa *Dione Juno* en el PMML.

Orden: Lepidoptera.

Familia: Nymphalidae.

Género: Dione.

Especie: Dione Juno.

Planta Hospedera: Pasiflora (Maracuyá).

✓ *Callicore Pitheas*



Figura 22. Foto de mariposa *Callicore Pitheas* en el PMML.

Orden: Lepidoptera.

Familia: Nymphalidae.

Género: Callicore.

Especie: Callicore Phiteas.

Planta Hospedera: Sapindácea Serjania.

✓ *Mechanitis Polymnia*



Figura 23. Foto de mariposa *Mechanitis Polymnia* en el PMML.

Orden: Lepidoptera.

Familia: Nymphalidae.

Género: Mechanitis.

Especie: Mechanitis Polymnia.

Planta Hospedera: Solanácea.

✓ *Colobura Dirce*



Figura 24. Foto de mariposa *Colobura Dirce* en el PMML.

Orden: Lepidoptera.

Familia: Nymphalidae.

Género: Colobura.

Especie: Colobura Dirce.

Planta Hospedera: Urticácea Yarumo.

Estas especies se encuentran actualmente en estudio activo con características ecológicas, morfológicas y de comportamiento ya identificadas por el grupo de investigación CIAM en el PMML y por esta razón fueron las especies seleccionadas para ingresarlas al sistema utilizando el módulo de registro **Figura 25**.

MENÚ MARIPOSAS

Methona Confusa

ESPECIE Methona Confusa

INFORMACIÓN

FAMILIA: Nymphalidae.
PLANTA HOSPEDERA: Francesina.
GENERO: Methona.
CICLO DE VIDA: 31 dias.

REGISTRAR MARIPOSA EDITAR MARIPOSA BORRAR MARIPOSA

Figura 25. Interfaz administración de mariposas REIDMAR v1.0.

La segunda fase correspondió a la digitalización de veinte fotografías por cada ejemplar en estudio, pasándolas de la cámara digital Samsung ES95, utilizada para realizar estas tomas en el mariposario del PMML, al computador **Figura 26** y luego cargándolas al sistema mediante un selector de archivos que sólo acepta formatos de imagen JPEG/JPG y PNG **Figura 27** con las propiedades de tamaño por defecto de la cámara.



Figura 26. Segunda fase metodológica.

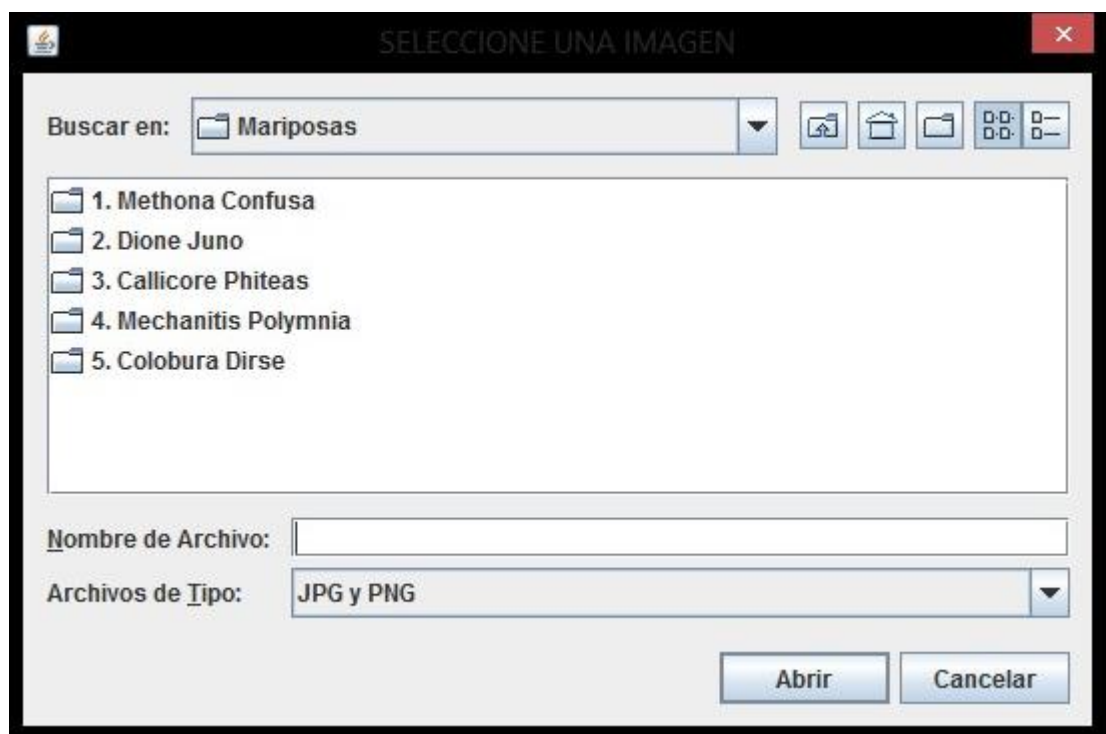


Figura 27. Selector de imágenes REIDMAR v1.0.

La tercera fase, en términos generales, correspondió al diagrama representado en la **Figura 28**, donde recibía la fotografía digitalizada de cada mariposa y después de la aplicación de técnicas de procesamiento de imágenes, **Figura 29**, devolvía un arreglo bidimensional o mapa de bits con un tamaño de 200x200 píxeles.



Figura 28. Tercera fase metodológica.



Figura 29. Diagrama detallado para el procesamiento de las imágenes.

El módulo para el procesamiento de imágenes recibió las imágenes y una a una les aplicó las técnicas de procesamiento de imágenes que se muestran en la **Figura 29**.

A continuación se muestra el proceso paso a paso que lleva de tener la imagen digital a convertirla en el arreglo bidimensional, para ello se utilizó la foto de la mariposa *Methona Confusa*.

El suavizado que “es una de las operaciones de convolución, se basa en que un píxel destino es el promedio de su correspondiente píxel origen con sus píxeles vecinos,

este tipo de filtrado es útil para la reducción del ruido” (Kono, n.d.), el resultado mostrado en la **Figura 30** se logró aplicando la **Ecuación 1** y la máscara de suavizado presentada en la **Figura 4**.

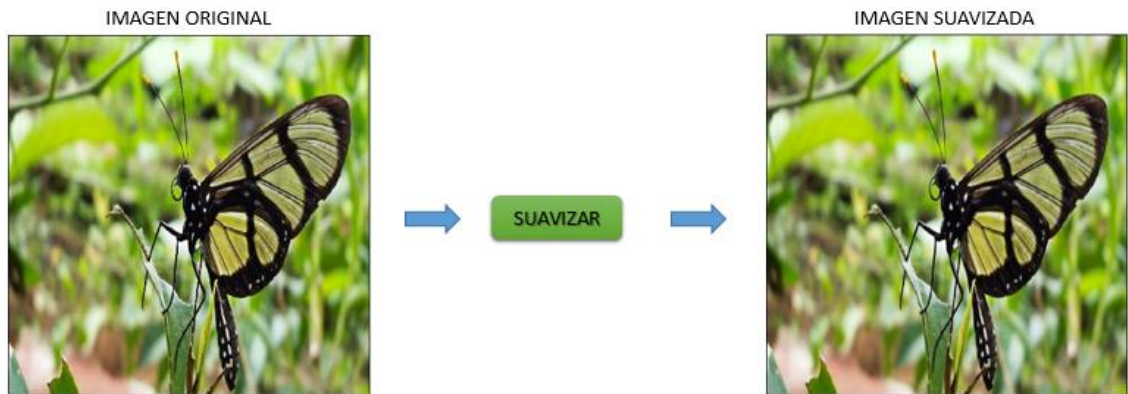


Figura 30. Técnica de suavizado aplicada.

La escala de grises que convirtió la característica del color de las imágenes a formato monocromático (Blanco/Negro) se consiguió con la implementación de la **Ecuación 2** y el resultado se puede apreciar en la **Figura 31**.



Figura 31. Técnica escala grises aplicada.

En la **Figura 32** se evidencia la ejecución de la técnica para detección de bordes resaltando los cambios en el color de la imagen utilizando la **Ecuación 1** junto con la máscara detecta bordes de la **Figura 4**.

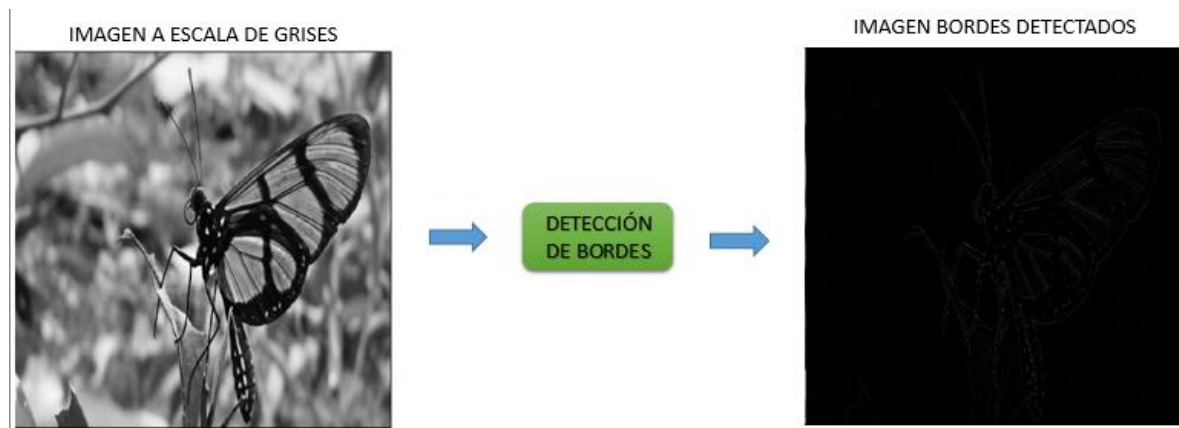


Figura 32. Técnica detección de bordes aplicada.

La transformación o escalamiento se usó para ajustar y estandarizar el tamaño de las fotografías dentro del sistema a un tamaño de 200x200 píxeles, que fue la mínima medida encontrada para evitar perder características fundamentales de las imágenes, manejando las **Ecuaciones 5 y 6**, la **Figura 33** expone el producto de aplicar este proceso.

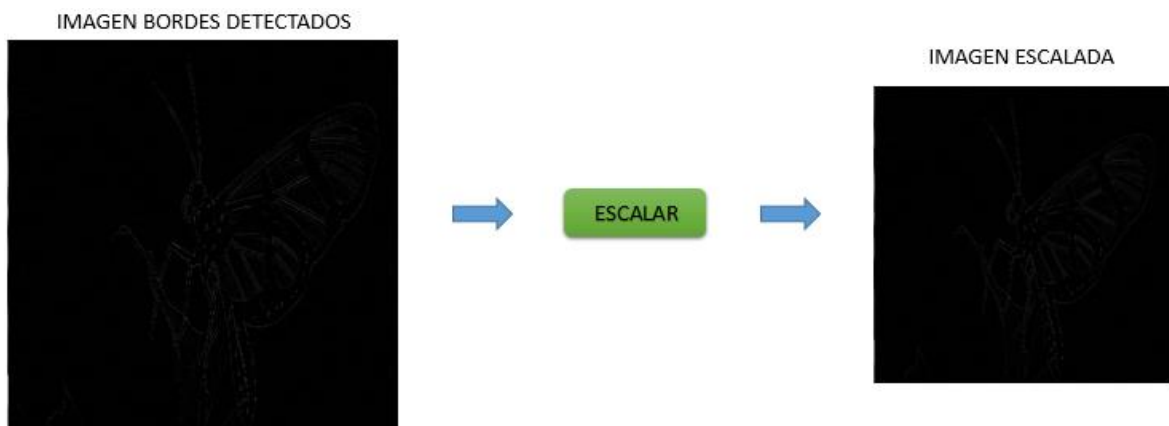


Figura 33. Técnica de transformación aplicada.

La binarización permitió separar la mariposa del fondo de la imagen haciendo uso de las **Ecuaciones 6, 7 y 8**, el resultado se observa en la **Figura 34**.

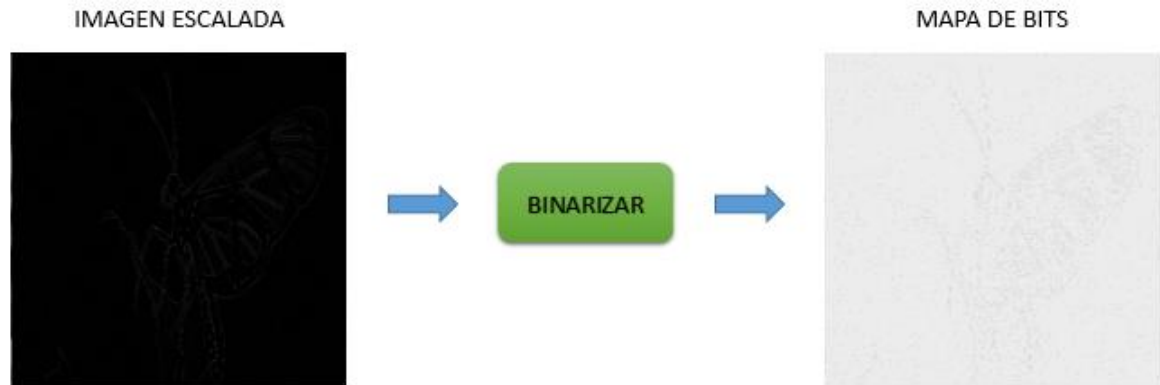


Figura 34. Técnica de binarización aplicada.

La cuarta fase, representada por la **Figura 35**, está determinada por la entrada (el mapa de bits 200x200 px), el proceso de reconocimiento de patrones, el proceso de entrenamiento y la salida que se una tabla de características donde cada especie está amarrada a un identificador único en el sistema.



Figura 35. Cuarta Fase Metodológica.

Para el reconocimiento de patrones, se creó una matriz 200x200 de índices que luego se multiplicó por la matriz binaria de la imagen también de 200x200, seguidamente, de la matriz resultante del producto, se calculó la suma tanto por filas como por columnas y estos 400 resultados se pasaron a un *array* o vector que se considera el patrón de entrenamiento para la RNA y el vector de observación para el HMM, este se convierte en el vector característico de cada imagen de mariposa. La **Figura 36** muestra el diagrama de esta fase.

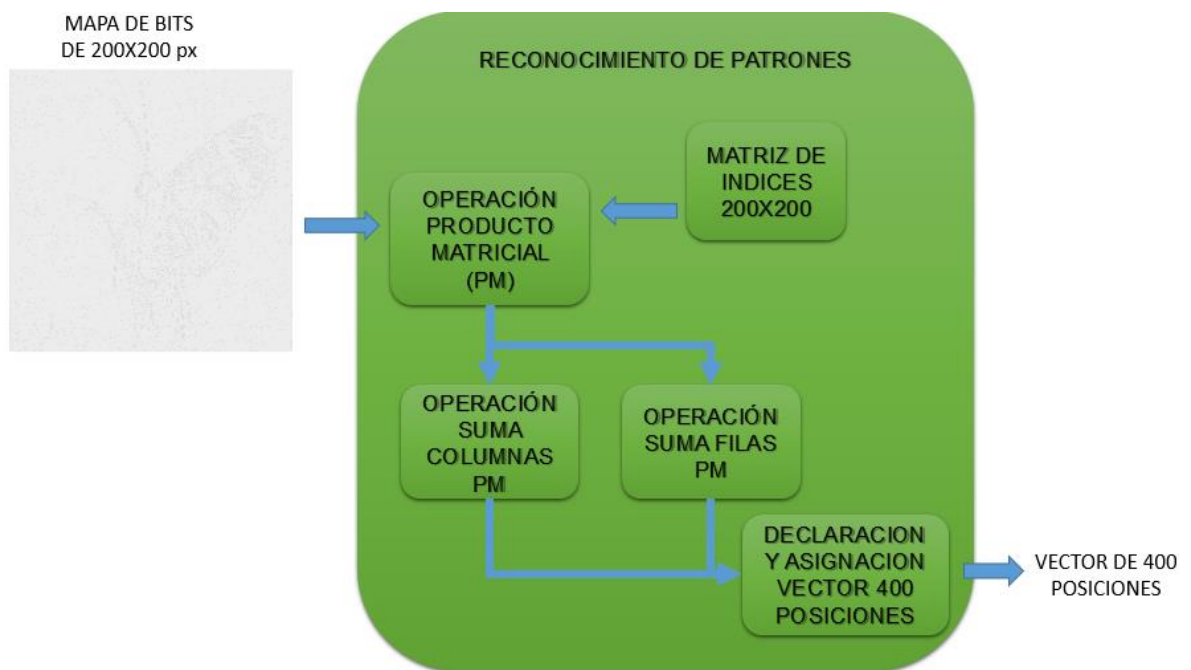


Figura 36. Diagrama reconocimiento de patrones.

La segunda parte de esta fase corresponde al entrenamiento de la RNA **Figura 38** y el HMM **Figura 39** explicado en el siguiente apartado, después de finalizado este paso se procedió a la fase de validación desde la vista del administrador **Figura 40**.

3.2. ENTRENAMIENTO DEL SISTEMA

Como se explicó desde un principio esta herramienta cuenta con la implementación de dos métodos de inteligencia artificial, el primero de ellos se trata de RNA en específico Backpropagation cuya arquitectura definida se presenta en la **Figura 37**, mostrando la entrada como el vector característico de 400 posiciones que se obtuvo del procesamiento de la imagen y el reconocimiento de patrones, la capa oculta de la RNA definida por 55 neuronas de procesamiento, cada una con función de activación sigmoidea y una capa de salida con tres neuronas correspondientes a la tabla característica de la especie **Figura 35**.

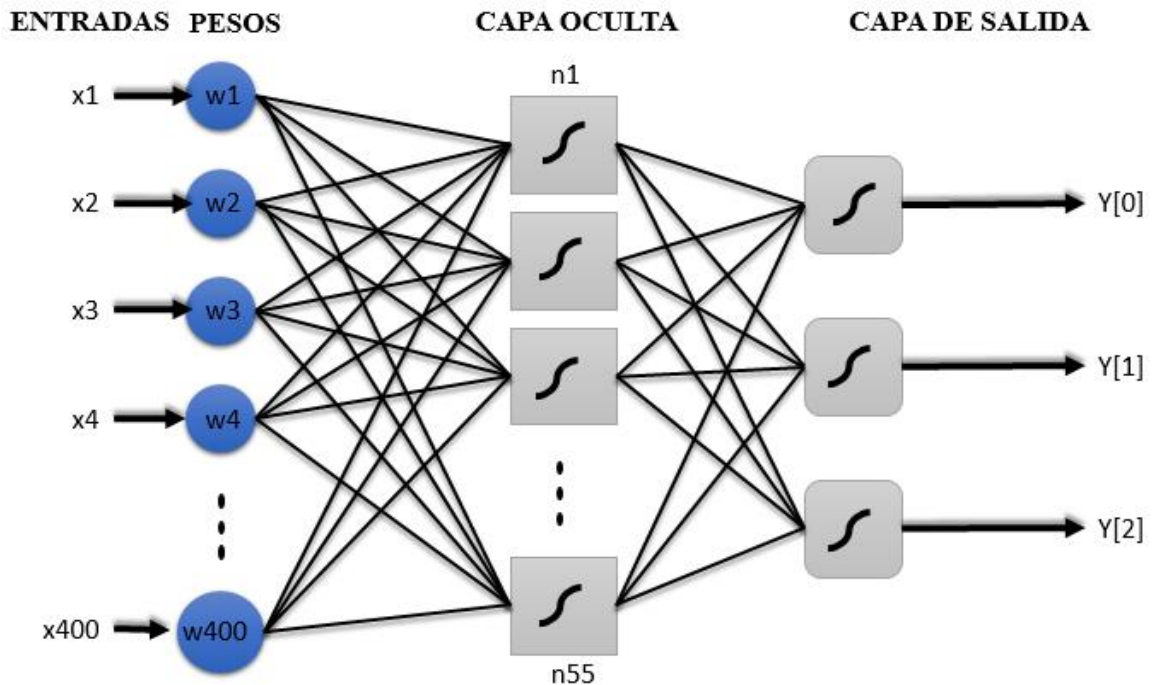


Figura 37. Diagrama entrenamiento RNA.

Con el modelo de RNA definido, se pasó a la etapa de aprendizaje utilizando el algoritmo de entrenamiento Backpropagation definido por (Hilera, Martínez, 1994) y las **Ecuaciones 10 – 19**, la **Figura 38** presenta el flujograma de entrenamiento para la RNA.

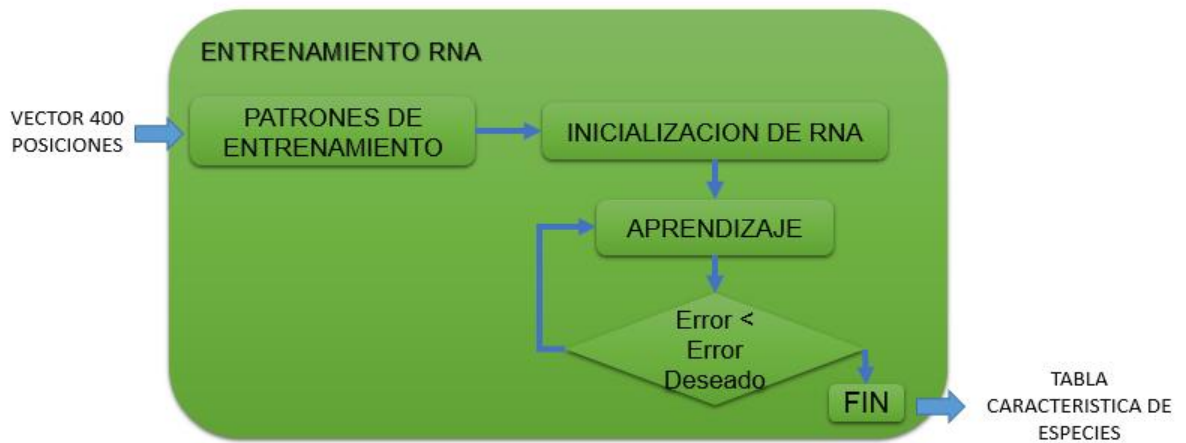


Figura 38. Diagrama entrenamiento RNA.

Posteriormente, se entrenó el HMM siguiendo el lineamiento planteado por (Jurafsky & Martin, 2017) con el algoritmo de aprendizaje Forward-Backward y aplicando las **Ecuaciones 20 – 24**. La arquitectura HMM que se planteó para el sistema consta de un modelo M por especie de mariposa, es decir cinco modelos en total, con dos estados ocultos representados por las palabras “SI” y “No” indicando que la mariposa pertenece o no al modelo **Figura 39**.

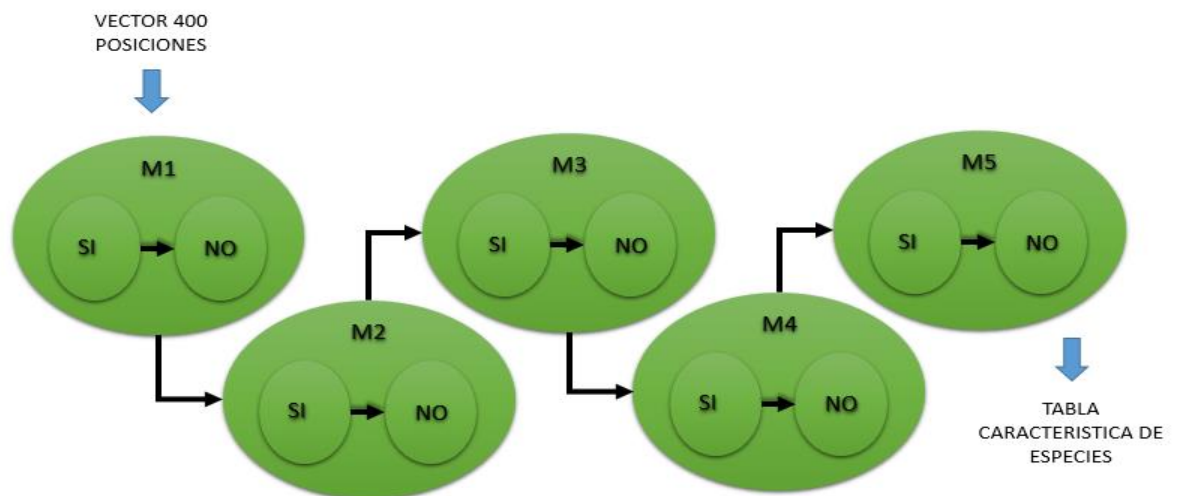


Figura 39. Diagrama entrenamiento HMM.

3.3. VALIDACIÓN DE RESULTADOS

En la quinta fase se desarrolló un módulo de validación **Figura 40**, que se ejecuta después de terminado el entrenamiento, el cual consistió en presentarle imágenes iguales, similares o diferentes a las utilizadas en la fase de aprendizaje obteniendo una respuesta del sistema y con ayuda de la experticia de un investigador del PMML se validó si la respuesta del sistema era correcta o se hacía necesario un nuevo entrenamiento de la herramienta.



Figura 40. Interfaz validación administrador REIDMAR v1.0.

Finalmente, utilizando los resultados provenientes de esta fase se determinó cuál de los métodos inteligentes en cuestión (RNA - HMM) responde eficientemente frente a la identificación de patrones y reconocimiento de las mariposas en estudio.

IV. ESTIMACION DE RECURSOS

| Rubros | Solicitado en efectivo a UDEC | Contrapartida en especie | | |
|--------------------|-------------------------------|--------------------------|----------------------|--------------------|
| | | UDEC | Otras Entidades | Total |
| PERSONAL | \$0 | Docente Investigador | - | \$0 |
| EQUIPOS | \$0 | | Compra de Computador | \$1'500.000 |
| MATERIALES INSUMOS | \$0 | - | - | \$0 |
| SERV. TECN. | \$0 | - | - | \$0 |
| VIAJES | \$500.000 | - | - | \$500.000 |
| OTROS | \$0 | - | - | \$0 |
| TOTALES | \$500.000 | \$0 | \$1'500.000 | \$2'000.000 |

Tabla 12. Recursos del Proyecto

V. RESULTADOS

Se utilizó un total de 100 patrones/observaciones de entrenamiento, es decir, 20 por cada una de las especies. Para las pruebas, se tomaron un total de 120 muestras por las cinco especies de mariposas, es decir, 24 fotografías por cada una. Obteniendo como resultado un 67,5% de acierto global del sistema, promediando las respuestas de los dos métodos RNA – HMM, para la identificación de las cinco especies.

En la **Tabla 13** se indican los resultados del reconocimiento de la RNA en la fase de verificación y en la **Tabla 14** se expresan los resultados del HMM, teniendo en cuenta que el 100% por especie corresponde a las 24 imágenes de prueba.

| | Porcentaje de Acierto RNA |
|----------------------------|------------------------------|
| <i>Methona Confusa</i> | 50% |
| <i>Dione Juno</i> | 67% |
| <i>Callicore Pitheas</i> | 42% |
| <i>Mechanitis Polymnia</i> | 67% |
| <i>Colobura Dirce</i> | 42% |

Tabla 13. Porcentajes de acierto RNA por especie de Mariposa

| | Porcentaje de Acierto HMM |
|----------------------------|--------------------------------------|
| <i>Methona Confusa</i> | 100% |
| <i>Dione Juno</i> | 83% |
| <i>Callicore Pitheas</i> | 75% |
| <i>Mechanitis Polymnia</i> | 83% |
| <i>Colobura Dirce</i> | 67% |

Tabla 14. Porcentajes de acierto HMM por especie de Mariposa

Con estos resultados, fácilmente se calculó y determinó cuál de los dos métodos de inteligencia artificial respondió mejor al problema de reconocimiento de las cinco especies de mariposas que habitan el PMML. Para el caso de la RNA su acierto total en porcentaje fue del 54% mientras que con respecto al HMM se encontró un porcentaje del 82%, para ambos casos, el 100% pertenece a las 120 muestras que se utilizaron para la validación del sistema REIDMAR v.1.0.

Se encontró un error global del 32,5% que puede deberse a varias razones como la ubicación y forma de la mariposa, la similitud frente a las otras especies, calidad de la imagen (iluminación – traslape – oclusión), entre otras.

5.1. CAPTURAS DE PANTALLA REIDMAR v.1.0



Figura 41. Presentación REIDMAR v1.0

MENÚ PRINCIPAL



ADMINISTRADOR



USUARIO GENERAL



ACERCA DE



SALIR

Figura 42. Menú Principal REIDMAR v1.0

ACERCA DE

DESARROLLO DE PROTOTIPO DE SISTEMA INTELIGENTE
PARA REGISTRO E IDENTIFICACIÓN DE MARIPOSAS
EN EL PARQUE METROPOLITANO MARIA LUCIA
(VILLAVICENCIO) BASADO EN HMM Y RNA.



VERSION BETA



AUTORES

JAIME EDUARDO ANDRADE RAMÍREZ (UDEC)
LUZ DARY PEDRAZA (UNIMETA)
OSCAR JOBANY GÓMEZ OCHOA (UDEC)

SEMILLERO DE ROBÓTICA DE FACATATIVÁ
(SERFAC)
GRUPO DE INVESTIGACIÓN DE FACATATIVÁ
(GISTFA)

CENTRO DE INVESTIGACIONES AMBIENTALES
JOSÉ ANTONIO CANDAMO [CIAM]



UNIVERSIDAD DE CUNDINAMARCA
CORPORACIÓN UNIVERSITARIA DEL META
COLOMBIA, 2017



Figura 43. Acerca de REIDMAR v1.0

MENÚ USUARIO GENERAL

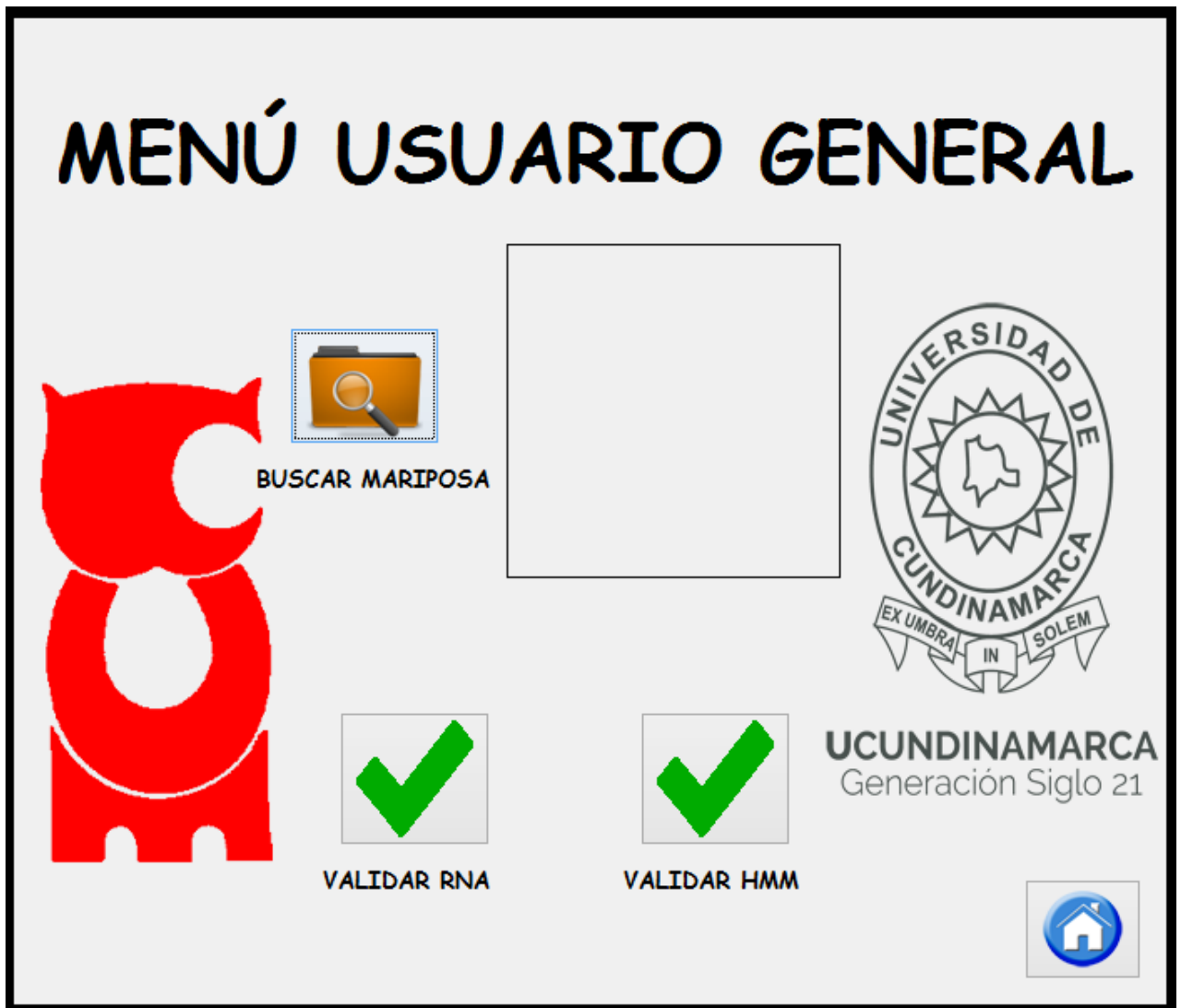


Figura 44. Menú Usuario General REIDMAR v1.0

ACCESO ADMINISTRADOR

Usuario:

Contraseña:



INGRESAR



LIMPIAR



Figura 45. Login Administrador REIDMAR v1.0

MENÚ ADMINISTRADOR



BUSCAR IMAGEN



GESTIONAR MARIPOSAS

Seleccione Mariposa... ▾



ENTRENAR SISTEMA



VALIDAR SISTEMA



Figura 46. Menú Administrador REIDMAR v1.0

MENÚ MARIPOSAS

Seleccione Mariposa... ▾

ESPECIE

INFORMACIÓN

 REGISTRAR MARIPOSA

 EDITAR MARIPOSA

 BORRAR MARIPOSA



Figura 47. Menú Mariposas REIDMAR v1.0

VALIDACIÓN PARA ADMIN



Figura 48. Menú Validación del Administrador REIDMAR v1.0

VI. CONCLUSIONES Y RECOMENDACIONES

Las técnicas para el procesamiento de las imágenes implementadas en este sistema son buenas ya que permiten obtener imágenes adecuadas minimizando características de tamaño y color pero cuidando la calidad de la imagen para no perder información relevante.

La imagen se limita a un formato monocromático para facilitar su manejo dentro de este sistema inteligente pero para lograr una extracción de características de las mariposas más específica es necesario manejar las imágenes a color.

Tanto la RNA Backpropagation como el HMM Forward-Backward implementados en este sistema funcionan bien para lograr el objetivo del reconocimiento de patrones. Pero en comparación el HMM logra ser más efectivo con un 82% de acierto global en el reconocimiento frente a un 54% de la RNA, ambos con 100 patrones/observaciones de entrenamiento y 60 imágenes de prueba.

En este documento, el almacenamiento de los datos se hace en una base de datos relacional MySQL. Este proceso puede mejorar teniendo en cuenta el desarrollo tecnológico actual y la vitalidad de mantener la información segura, integra y disponible en todo momento; implementando bases de datos no relacionales, para aumentar la velocidad y capacidad de procesamiento de los datos.

VII. BIBLIOGRAFIA

- Andrade C, M. G. (2022). Medio ambiente Estado Del Conocimiento De La Interacción Ciencia-Política. *Revista de La Académica Colombiana de Ciencias*, 35(137), 491–507. Recuperado de <http://www.scielo.org.co/pdf/racefn/v35n137/v35n137a08.pdf>
- Martha Rubio, Fabio Jiménez (2015). *Mariposas del Darién*. Editorial LEGIS S.A, Sapzurro
- Minorta-C, V. & J.O. Rangel-Ch (2014). Las riquezas y la diversidad de las plantas con flores de la Orinoquia colombiana. En: J.O. Rangel-Ch. (ed.) *Colombia Diversidad Biótica XIV. La Región Orinoquia de Colombia*: 237-418. Universidad Nacional de Colombia – Instituto de Ciencias Naturales. Bogotá.
- Constantino LM. (1997). Conocimiento de los ciclos de vida y plantas hospederas de lepidópteros diurnos de Colombia como estrategia para el manejo, uso y conservación de poblaciones silvestres. *Acontecer Entomológicos*, 57-81.
- Francesconi, W., Nair, P. K. R, Levey, D. J., Daniels, J., & Cullen, L. (2013). Butterfly distribution in fragmented landscapes containing agroforestry practices in Southeastern Brazil. *Agroforestry System*, 87(6), 1321-1338. <https://doi.org/10.1007/s10457-013-9640-v>
- Ray Liu, K. J, (2002). *Pattern Recognition and Image Preprocessing*.
- Dębska, B., & Guzowska-Świder, B. (2011). Application of artificial neural network in food classification. *Analytica Chimica Acta*, 705(1), 283–291. <https://doi.org/10.1016/j.aca.2011.06.033>
- Jurafsky, D., & Martin, J. (2017). Hidden Markov Models. *Speech and Language Processing*, (Chapter 20), 21. [https://doi.org/10.1016/S0959-440X\(96\)80056-X](https://doi.org/10.1016/S0959-440X(96)80056-X)
- Górska-Horczyk, E., Horczyk, M., Guzek, D., Wojtasik-Kalinowska, I., & Wierzbicka, A. (2017). Chromatographic fingerprints supported by artificial neural network for differentiation of fresh and frozen pork. *Food Control*, 73, 237–244. <https://doi.org/10.1016/j.foodcont.2016.08.010>

- Murru, N., & Rossini, R. (2016). Neurocomputing A Bayesian approach for initialization of weights in backpropagation neural net with application to character recognition. *Neurocomputing*, 193, 92–105. <https://doi.org/10.1016/j.neucom.2016.01.063>
- Kouakou, A. K., Bagui, O. K., Agneroh, T. A., Soro, A. P., & Zoueu, J. T. (2016). Cucumber mosaic virus detection by artificial neural network using multispectral and multimodal imagery. *Optik - International Journal for Light and Electron Optics*, 127(23), 11250–11257. <https://doi.org/10.1016/j.ijleo.2016.09.035>
- Fan, F. H., Ma, Q., Ge, J., Peng, Q. Y., Riley, W. W., & Tang, S. Z. (2013). Prediction of texture characteristics from extrusion food surface images using a computer vision system and artificial neural networks. *Journal of Food Engineering*, 118(4), 426–433. <https://doi.org/10.1016/j.jfoodeng.2013.04.015>
- Olawoyin, R. (2016). Chemosphere Application of backpropagation artificial neural network prediction model for the PAH bioremediation of polluted soil. *Chemosphere*, 161, 145–150. <https://doi.org/10.1016/j.chemosphere.2016.07.003>
- Zhou, M., Zhang, X., Yin, F., & Liu, C. (2016). Discriminative quadratic feature learning for handwritten Chinese character recognition. *Pattern Recognition*, 49, 7–18. <https://doi.org/10.1016/j.patcog.2015.07.007>
- Surinta, O., Karaaba, M. F., Schomaker, L. R. B., & Wiering, M. A. (2015). Engineering Applications of Arti fi cial Intelligence Recognition of handwritten characters using local gradient feature descriptors. *Engineering Applications of Artificial Intelligence*, 45, 405–414. <https://doi.org/10.1016/j.engappai.2015.07.017>
- Mofavvaz, S., Sohrabi, M. R., & Nezamzadeh-Ejhieh, A. (2017). New model for prediction binary mixture of antihistamine decongestant using artificial neural networks and least squares support vector machine by spectrophotometry

- method. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 182, 105–115. <https://doi.org/10.1016/j.saa.2017.04.001>
- Lopes, A. T., de Aguiar, E., De Souza, A. F., & Oliveira-Santos, T. (2017). Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order. *Pattern Recognition*, 61, 610–628. <https://doi.org/10.1016/j.patcog.2016.07.026>
- Kaya, Y., Kayci, L., & Uyar, M. (2015). Automatic identification of butterfly species based on local binary patterns and artificial neural network. *Applied Soft Computing*, 28(2015), 132–137. <https://doi.org/10.1016/j.asoc.2014.11.046>
- Liisberg, J., Møller, J. K., Bloem, H., Cipriano, J., Mor, G., & Madsen, H. (2016). Hidden Markov Models for indirect classification of occupant behaviour. *Sustainable Cities and Society*, 27, 83–98. <https://doi.org/10.1016/j.scs.2016.07.001>
- Bhatti, S. M., Khan, M. S., Wuth, J., Huenupan, F., Curilem, M., Franco, L., & Yoma, N. B. (2016). Automatic detection of volcano-seismic events by modeling state and event duration in hidden Markov models. *Journal of Volcanology and Geothermal Research*, 324, 134–143. <https://doi.org/10.1016/j.jvolgeores.2016.05.015>
- Premaratne, P., Yang, S., Vial, P., & Ifthikar, Z. (2017). Centroid tracking based dynamic hand gesture recognition using discrete Hidden Markov Models. *Neurocomputing*, 228(June 2016), 79–83. <https://doi.org/10.1016/j.neucom.2016.06.075>
- Jimenez-Fernandez, V. M., Vazquez-Leal, H., Filobello-Nino, U. A., Jimenez-Fernandez, M., Morales-Mendoza, L. J., & Gonzalez-Lee, M. (2016). Exploring the use of two-dimensional piecewise-linear functions as an alternative model for representing and processing grayscale-images. *Journal of Applied Research and Technology*
- SCRUMstudy. (2017). *Una guía para el cuerpo de conocimiento de Scrum*. Retrieved from <http://www.scrumstudy.com/SBOK/SCRUMstudy-SBOK-Guide-3rd-edition-spanish.pdf>

- Melo, M. D. (1970). Técnicas en el Procesamiento Digital de Imágenes: Una Constante Exploración de Nuevos Campos, 10–19.
- Huang, X., Jack, M., & Ariki, Y. (1990). Hidden Markov Models for Speech Recognition, *33*(3), 251–272.
- Manuel, C., & González, T. (2002). DEPARTAMENTO DE SEÑALES Y COMUNICACIONES TESIS DOCTORAL RECONOCIMIENTO DE FORMAS MANUSCRITAS CON MODELOS OCULTOS DE MARKOV CARLOS MANUEL TRAVIESO GONZÁLEZ Las Palmas de Gran Canaria, Mayo de 2002.
- Hoyos, M. H., & Hoyos, M. H. (2004). Procesamiento y Análisis de Imágenes Digitales (PAID) Que es PAID ?
- Rojas, R., Sarmiento, S., Lawrence, W., & Escobar, V. (2007). Reconocimiento de patrones en un arreglo sensorico usando redes neuronales, *17*, 95–111. Retrieved from <http://dialnet.unirioja.es/servlet/oaiart?codigo=2387888>
- A PROBABILISTIC PRINCIPAL COMPONENT ANALYSIS BASED HIDDEN MARKOV MODEL FOR AUDIO-VISUAL SPEECH RECOGNITION Zhanyu Ma and Arne Leijon Sound and Image Processing Lab , KTH - Royal Institute of Technology. (2008). *Image Processing*, 2170–2173.
- Sáez, A. P. (2009). Reconocimiento De Imágenes a Través De Su Contenido, 248. Retrieved from <http://www.iit.comillas.edu/pfc/resumenes/4a4488b9a17fe.pdf>
- Moler, E., & Pastore, J. I. (2010). Técnicas De Procesamiento Digital De Imágenes : Una Aplicación Para Identificación De Personas a Través, *XXIX*, 15–18.
- Cattaneo, C. A., Larcher, L. I., Ruggeri, A. I., Herrera, A. C., & BIASONI, E. M. (2011). Métodos De Umbralización De Imágenes Digitales Basados En Entropía De Shannon Y Otros, *XXX*, 2785–2805.
- Im, D. E., Digitales, G., Algoritmia, S. U., Herramienta, C., Ilustraci, L. A., & Magro, G. R. (2013). Binarización De Imágenes Digitales Y Su Algoritmia Como Herramienta Aplicada a La Ilustración Entomológica, *53*, 443–464.

- Andrade, J., Gomez, O., & Rodríguez, C. (2016). PROTOTIPO DE SISTEMA INTELIGENTE PARA LA IDENTIFICACION Y REGISTRO DE MARIPOSAS NATIVAS DE LA REGION ORINOQUIA MEDIANTE REDES NEURONALES ARTIFICIALES IMPLEMENTADAS EN JAVA, 1–6.
- Jaime, I., Andrade, E., Cundinamarca, U. De, Luz, L., Pedraza, D., Universitaria, C, Moreno, R. (2017). Inteligencia Artificial como Herramienta para Reconocimiento de Mariposas en Villavicencio (Caso de Estudio Parque Metropolitano María Lucia), 2017(Eaci), 1–6.
- Universitaria, C. (2018). Registro e Identificación de Mariposas en Villavicencio-Colombia Aplicando Inteligencia Artificial (Caso de Estudio Parque Metropolitano María Lucia), 21–25.
- SING-TZE, Bow; Pattern Recognition Image Preprocessing (2nd Edition). Marcel Dekker Incorporated Editors, 2002.
- Seijas, Leticia María; Reconocimiento de Dígitos Manuscritos Mediante Redes Neuronales: Una Técnica Híbrida.
- Marín Ortiz, Manuel; Procesamiento Digital de Imágenes. Universidad Autónoma de Puebla, Enero 2013
- Hernández, Nelson; Fundamentos de Procesamiento Digital de Imágenes. Revista de Tecnología, Vol.3, N°2, Julio-Diciembre 2004
- PHP (2015). Filtros de convolución. <http://www.rinconastur.com/php/php134.php>
- Giménez Palomares, F., Monsoriu Serrá, J. A., & Alemany Martínez, E. (2016). Aplicación de la convolución de matrices al filtrado de imágenes. *Modelling in Science Education and Learning*, 9(1), 97. <https://doi.org/10.4995/msel.2016.4524>
- Jimmy, M. S., & Cortés, A. (2011). TÉCNICAS ALTERNATIVAS PARA LA CONVERSIÓN DE IMÁGENES A COLOR A ESCALA DE GRISES EN EL TRATAMIENTO DIGITAL DE IMÁGENES Color images Conversion Alternatives Techniques to grayscale in Digital Image Processing, (47), 207–212.

- Melo, S. B. (n.d.). TRANSFORMACIONES GEOMÉTRICAS SOBRE IMÁGENES DIGITALES, 1–13.
- Chuk, T., Crookes, K., Hayward, W. G., Chan, A. B., & Hsiao, J. H. (2017). Hidden Markov model analysis reveals the advantage of analytic eye movement patterns in face recognition across cultures. *Cognition*, 169, 102–117. <https://doi.org/10.1016/j.cognition.2017.08.003>
- Acosta, X., Nieto, E., & Barahona, C. (2015). Metodología para la evaluación de calidad de los productos software de la Universidad de Cundinamarca. ENGI Revista Electrónica De La Facultad De Ingeniería.
- Rumbaugh, J, Jacobson, I, Booch, G. (2000). El Lenguaje Unificado de Modelado. Manual de Referencia.
- Kono, M. D. (n.d.). Imágenes en Java.