

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



**Diseño e Implementación de un Sistema de Monitorización de
Posibles Derrumbes a través de Visión por Computador**

Trabajo de graduación presentado por José Miguel Castañeda Chavarría
y Silvio Andrés Orozco Vizquerra para optar al grado académico de
Licenciados en Ingeniería en Ciencias de la Computación y Tecnologías
de la Información

Guatemala,

2022

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



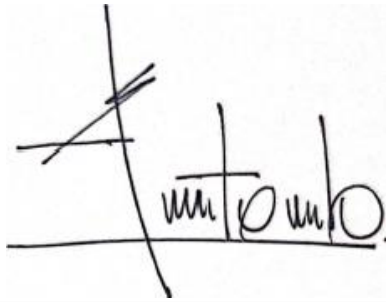
**Diseño e Implementación de un Sistema de Monitorización de
Posibles Derrumbes a través de Visión por Computador**

Trabajo de graduación presentado por José Miguel Castañeda Chavarría
y Silvio Andrés Orozco Vizquerra para optar al grado académico de
Licenciados en Ingeniería en Ciencias de la Computación y Tecnologías
de la Información

Guatemala,

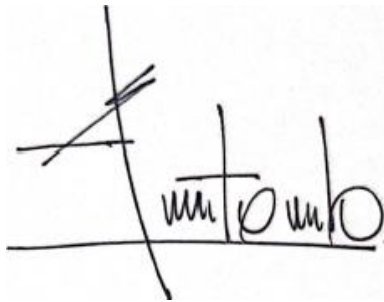
2022

Vo.Bo.:



(f) _____
Ing Danilo Antonio Rodríguez Cerón

Tribunal Examinador:



(f) _____
Ing Danilo Antonio Rodríguez Cerón



(f) _____
MSc Douglas Leonel Barrios González



(f) _____
Ing Leonel Guillen del Valle

Fecha de aprobación: Guatemala, 27 de enero de 2023.

En los últimos años, ha habido muchas incidencias de deslizamientos en Guatemala. Esto a provocado daño en las carreteras, las cuales cuestan una gran cantidad de dinero al país y generan tráfico y obstrucciones viales. Además, muchas familias, sobre todo familias de bajos recursos, viven en asentamientos cerca de áreas susceptibles a estos fenómenos. Por lo que el objetivo de este proyecto es poder brindar una herramienta que sirva para mejorar la planificación de viviendas y carreteras en el país y disminuir la cantidad de personas que son afectadas por los deslizamientos.

Este trabajo se lo dedicamos a Dios que nos ha brindado la oportunidad de poder estudiar y aprender en una buena universidad.

Queremos agradecer a nuestros padres, Silvio Orozco y Ana Beatriz Vizquerra, Mynor Castañeda y Brenda de Castañeda que han trabajado arduamente para siempre darnos las herramientas necesarias para superarnos como persona y que nos han enseñado sus valores para formarnos como personas íntegras y conscientes de los demás; a nuestro asesor, ingeniero Danilo Rodríguez, por todo su apoyo y conocimiento durante este trabajo de graduación; al ingeniero Douglas Barrios, director de ingeniería en ciencias de la computación y tecnologías de la información, por su apoyo durante toda la carrera y su buena disposición para ayudarnos en cualquier momento; a CONRED, INSIVUMEH y MAGA por brindar información relevante de mapas que aportan a este proyecto; a nuestros profesores involucrados en nuestra formación académica, que durante estos años nos han aportado la pasión por las ciencias de la computación y nos han alentado a perseguir la excelencia; y a la Universidad del Valle de Guatemala y al colegio Liceo Javier, por ser nuestras casas de estudio en las que hemos formado nuestros valores, criterios y conocimientos y se nos ha enseñado a tener nuestro propio criterio, ser autodidactas y estar siempre abierto a aprender de cualquier persona.

Prefacio	III
Lista de figuras	IX
Lista de cuadros	XII
Resumen	XIII
Abstract	XIV
1. Introducción	1
2. Justificación	3
3. Objetivos	4
3.1. Objetivo general	4
3.2. Objetivos específicos	4
4. Marco teórico	5
4.1. Derrumbes	5
4.1.1. ¿Qué es un derrumbe?	5
4.1.2. Categorías de tipos de deslizamiento	6
4.1.2.1. Caída	6
4.1.2.2. Derribo	7
4.1.2.3. Deslizamiento por derrumbe rotacional	8
4.1.2.4. Deslizamiento de traslación	8
4.1.2.5. Corriente o flujo	9
4.1.3. Principios de estabilidad de taludes	10
4.1.4. Causas de deslizamientos	10
4.1.5. Factores causantes de deslizamientos de tierra	11
4.2. Importancia de evaluación de peligro de deslizamientos	12
4.2.1. Mapas de inventarios de deslizamientos	12
4.2.2. Mapas de susceptibilidad	13

4.2.2.1. Enfoques de mapeo de amenazas y susceptibilidad a deslizamientos	15
4.2.2.2. Validación de modelos de mapas de susceptibilidad	18
4.3. Sistema de información geográfica (<i>GIS</i>)	19
4.3.1. Funciones principales	19
4.3.1.1. Almacenamiento	19
4.3.1.2. Visualización	19
4.3.1.3. Consultas	19
4.3.1.4. Análisis	19
4.3.1.5. Toma de decisiones	20
4.3.1.6. Modelación	20
4.3.2. Tipos de archivos manejados en un <i>GIS</i>	20
4.3.2.1. GeoTiff (.tiff)	20
4.3.2.2. Modelo digital de elevación (<i>DEM</i>)	20
4.3.2.3. Ortomosaico	21
4.3.2.4. <i>Shapefile</i>	21
4.3.3. Almacenamiento de información en un <i>GIS</i>	21
4.3.4. <i>Digital Terrain Analysis (DTA)</i>	22
4.4. Factores espaciales para la evaluación de derrumbes	23
4.4.1. Factores ambientales	24
4.4.1.1. Factores geomorfológicos: <i>DEM: elevation y slope</i>	24
4.4.1.2. Factores geológicos litológicos	26
4.4.1.3. Factores de suelo e hidrogeológicos	27
4.4.1.4. Factores capacidad del uso del suelo	38
4.4.1.5. Factores fisiográficos	40
4.4.2. Factores disparadores	40
4.4.2.1. Factor precipitación	40
4.4.3. Inventario de deslizamientos (base de datos)	41
4.5. Vehículos aéreos no tripulados	41
4.5.1. Tipos de UAV	41
4.5.2. Nivel de automatización	42
4.5.3. Fuentes de energía	43
4.6. Modelos de <i>machine learning</i>	43
4.6.1. <i>Machine learning</i>	43
4.6.2. Preprocesamiento del <i>training set</i>	45
4.6.3. Algoritmos de clasificación	46
4.6.3.1. <i>Decision tree</i> y <i>Random forest</i>	46
4.6.3.2. <i>AdaBoost</i> , <i>XGBoost</i> y <i>LightGBM</i>	49
4.6.3.3. Métodos probabilísticos y estadísticos: <i>Logistic regression</i> y <i>Support vector machine</i>	51
4.6.3.4. Evaluación y validación del <i>performance</i> los algoritmos	54
4.6.3.5. Evaluación del modelo y validación	54
4.6.3.6. <i>Hyperparameter optimization</i>	54
4.7. Segmentación semántica de imágenes	55
4.7.1. Métricas para modelos de segmentación semántica de imágenes	55
4.7.2. Algoritmos para segmentación semántica de imágenes	56
4.7.2.1. U-Net	56
4.7.2.2. <i>Feature Pyramid Network (FPN)</i>	56

4.7.2.3. <i>Pyramid Scene Parsing Network (PSPNet)</i>	57
4.7.2.4. <i>Path Aggregation Network (PANet)</i>	57
4.7.2.5. <i>DeepLab V3+</i>	57
5. Metodología	58
5.1. Metodología general	58
5.2. Proceso X - Generación del modelo de datos	63
5.3. Proceso A - Generación de inventarios de deslizamientos en Guatemala	75
5.4. Proceso B - Generación dataset para entrenamiento de modelo de susceptibilidad a deslizamientos	78
5.5. Proceso C - Entrenamiento modelos de <i>machine learning</i> con dataset	85
5.6. Proceso D - Generación puntos de Guatemala a 1km de distancia e integración de datos ligados al modelo	88
5.7. Proceso E - Evaluación de susceptibilidad de derrumbes en toda Guatemala escala 1km	90
5.8. Proceso F: Implementación y validación del sistema de monitorización de posibles derrumbes	93
5.9. Proceso Z - Entrenamiento de modelo de segmentación semántica	94
6. Resultados	97
6.1. Totales de datos de inventarios de deslizamiento	97
6.2. Estructura del modelo	103
6.3. Datos geomorfológicos obtenidos	105
6.4. Resultados <i>ROC AUC Y accuracy</i> modelos entrenados	106
6.5. Importancia de variables de modelos	107
6.5.1. Importancias usando media	107
6.5.2. Importancias usando punto exacto	111
6.6. Validación de modelos con <i>Landslide Density</i> y <i>R index</i>	114
6.6.1. <i>Landslide Density</i> usando media	115
6.6.2. <i>Landslide Density</i> usando punto exacto	116
6.6.3. <i>Landslide R Index</i> usando media	118
6.6.4. <i>Landslide R Index</i> usando punto exacto	119
6.7. Mapas de susceptibilidad generados por los modelos	122
6.8. Resumen de resultados y aprobación de modelos	124
6.9. Plataforma de susceptibilidad de deslizamientos	126
6.10. Modelos de segmentación semántica de imágenes	129
6.10.1. Resultados de los diferentes modelos	129
6.10.2. Resultados del modelo final de segmentación semántica	129
7. Discusión de resultados	131
7.1. Sobre la generación del inventario y dataset de deslizamientos	131
7.2. Sobre los modelos de predicción	134
7.2.1. <i>Decision tree</i>	137
7.2.2. <i>Random forest</i>	138
7.2.3. <i>XGBoost</i>	138
7.2.4. <i>Logistic regression</i>	139
7.2.5. <i>Support vector machine</i>	139
7.3. Sobre los mapas de susceptibilidad y la plataforma	140

7.4. Sobre los modelos de segmentación semántica	141
8. Conclusiones	143
9. Recomendaciones	145
10. Bibliografía	147

Lista de figuras

1. Ejemplo de derrumbe de tierra en Guatemala en libramiento de Chimalte- nango	6
2. Ejemplo de caída de rocas en Guatemala en carretera de Chiquimula	7
3. Esquema de caída de roca	7
4. Esquema de un derribo	7
5. Esquema de un deslizamiento rotacional	8
6. Ejemplo de un derrumbe rotacional	8
7. Esquema de un deslizamiento de traslación	9
8. Ejemplo de un derrumbe de traslación en Baja Verapaz, Guatemala	9
9. Esquema de una corriente o flujo	9
10. Estados de estabilidad de un talud	10
11. Ejemplo de un mapa de inventario de deslizamientos	13
12. Mapa de susceptibilidad de derrumbes en Guatemala	14
13. Evolución del almacenamiento de datos de <i>gis</i>	22
14. Curvatura de perfil	23
15. Curvatura del plano	23
16. Factores importantes para la evaluación de susceptibilidad de derrumbes	24
17. Clasificación de textura de suelos	29
18. UAV de ala fija	42
19. UAV multirrotor	42
20. Diagrama de clasificación binaria de derrumbes	44
21. Diagrama de aprendizaje supervisado	45
22. ¿Cómo funciona <i>Random forest</i> ?	48
23. Diagrama del flujo del <i>boosting method</i>	50
24. <i>Logistic model</i>	52
25. Gráfica de <i>logistic regression</i>	53
26. IoU	56
27. Diagrama de metodología general	59
28. Diagrama de modelo de datos	60
29. Diagrama de la matriz de datos geomorfológicos obtenidos del terreno	63
30. Diagrama obtención de datos para modelo de susceptibilidad: proceso x	64
31. Diagrama de generación del modelo de datos	76

32. Diagrama generación de dataset para entrenamiento de modelo de susceptibilidad a deslizamientos: proceso b	79
33. Inventario de deslizamientos en <i>QGIS</i>	81
34. <i>Heatmap landslides raster</i> params en <i>QGIS</i>	82
35. <i>Heatmap landslides raster</i> generado en <i>qgis</i>	82
36. <i>Sample raster values</i> del <i>landslide heatmap</i> en <i>QGIS</i>	83
37. Dataset con <i>landslides</i> y <i>non landslides</i> en <i>QGIS</i>	84
38. Diagrama generación de dataset para entrenamiento de modelo de susceptibilidad a deslizamientos: proceso c	86
39. Diagrama generación de puntos de Guatemala a 1km de distancia e integración de datos ligados al modelo: proceso D	89
40. Diagrama de evaluación de susceptibilidad de derrumbes en toda Guatemala escala 1km: proceso E	91
41. Diagrama de implementación y validación del sistema de monitorización de posibles derrumbes: proceso f	93
42. Mapa de Calor de deslizamientos en Guatemala	98
43. Mapa de Slope en Guatemala utilizando <i>DEM</i>	99
44. Mapa de deslizamientos en Guatemala	100
45. Mapa de <i>training</i> de deslizamientos en guatemala	101
46. Mapa de <i>testing</i> de deslizamientos en Guatemala	102
47. Diagrama del modelo de datos	103
48. Diagrama comparativo de datos geomorfológicos obtenidos de un ortomosaico y una imagen satelital	105
49. Matrices de confusión para modelos usando media	106
50. Matrices de confusión para modelos usando punto exacto	107
51. Importancia/peso de variables para modelos usando media	107
52. Ejemplo para árboles de decisión con <i>max depth 3</i>	108
53. Importancia/peso de variables para modelos usando punto exacto	111
54. Ejemplo para árboles de decisión con <i>max depth 3</i>	112
55. Mapa de susceptibilidad para modelos usando media	122
56. Mapa de susceptibilidad para modelos usando punto exacto	123
57. <i>Roc auc curve scores</i> usando media	124
58. <i>Precision vs recall curve scores</i> usando media	125
59. <i>Roc auc curve scores</i> usando punto exacto	125
60. <i>Precision vs recall curve scores</i> usando punto exacto	126
61. Diagrama de diseño del sistema de monitorización de derrumbes	126
62. Imagen de plataforma 1	127
63. Imagen de plataforma 2	127
64. Imagen de plataforma 3	128
65. Diseño de la base de datos	128
66. Clases de la segmentación semántica	129
67. Segmentación semántica con <i>validation data</i>	129
68. Segmentación semántica con <i>test data</i>	130
69. Segmentación semántica de ortomosaico	130

Lista de cuadros

1. Tipos de factores causantes de deslizamientos	11
2. Factores causantes de deslizamientos	12
3. Métodos cualitativos	16
4. Métodos cuantitativos	17
5. Métricas de evaluación de modelos de susceptibilidad	18
6. Factores geomorfológicos importantes para el análisis de derrumbes	25
7. Resumen de características catálogos disponibles en Google Earth de imágenes satelitales y <i>DEMs</i>	26
8. Resumen de propiedades litológicas obtenidas del mapa geológico de Guate- mala	27
9. Clasificación de las partículas del suelo	30
10. Resumen de propiedades taxonómicas de suelos obtenidas del mapa taxonó- mico de suelos Guatemala	31
11. Resumen de propiedades de reconocimiento de suelos Simmons obtenidas del mapa de reconocimiento de suelos Guatemala	31
12. Resumen de propiedades uso de suelo obtenidas del mapa de uso de suelos Guatemala	39
13. Resumen de propiedades fisiográficas obtenidas del mapa de fisiografía de Guatemala	40
14. Parámetros importantes de <i>decisiontreeclassifier</i> en librería scikit-learn de python	47
15. Parámetros importantes de <i>randomforestclassifier</i> en librería scikit-learn de python	48
16. Parámetros importantes de <i>adaboostclassifier</i> en librería scikit-learn de python	50
17. Parámetros importantes de <i>xgbclassifier</i> en librería <i>xgboost</i> de python	51
18. Parámetros importantes de <i>lgbclassifier</i> en librería <i>lightgbm</i> de python	51
19. Parámetros importantes de <i>supportvectorclassification</i> en librería <i>sklearn</i> de python	53
20. Modelo de datos creado para alimentar el modelo de susceptibilidad	60
21. Herramientas tecnológicas utilizadas en el proceso x	64
22. Librerías tecnológicas utilizadas en el proceso x	65

23. Archivos utilizados en el proceso x	66
24. Herramientas tecnológicas utilizadas en el proceso a	76
25. Librerías tecnológicas utilizadas en el proceso a	77
26. Archivos utilizados en el proceso a	77
27. Herramientas tecnológicas utilizadas en el proceso b	78
28. Librerías tecnológicas utilizadas en el proceso b	78
29. Archivos utilizados en el proceso b	80
30. Herramientas tecnológicas utilizadas en el proceso c	85
31. Librerías tecnológicas utilizadas en el proceso c	85
32. Archivos utilizados en el proceso c	87
33. Herramientas tecnológicas utilizadas en el proceso d	88
34. Librerías tecnológicas utilizadas en el proceso d	88
35. Archivos utilizados en el proceso d	89
36. Herramientas tecnológicas utilizadas en el proceso e	91
37. Librerías tecnológicas utilizadas en el proceso e	92
38. Archivos utilizados en el proceso e	92
39. Herramientas tecnológicas utilizadas en el proceso z	94
40. Librerías tecnológicas utilizadas en el proceso z	95
41. Archivos utilizados en el proceso z	96
42. Total de datos inventario deslizamientos	97
43. <i>Accuracy</i> y <i>roc auc score</i> para <i>train test</i> usando media	106
44. <i>Accuracy</i> y <i>roc auc score</i> para <i>train test</i> usando punto exacto	106
45. Importancia de variables para árboles de decisión usando media	108
46. Importancia de variables para <i>random forest</i> usando media	109
47. Importancia de variables para <i>xgboost</i> usando media	109
48. Coeficientes de variables para <i>logisticregression</i> usando media	110
49. Coeficientes de variables para <i>supportvectormachine</i> usando media	110
50. Importancia de variables para árboles de decisión usando punto exacto	111
51. Importancia de variables para <i>random forest</i> usando punto exacto	112
52. Importancia de variables para <i>xgboost</i> usando punto exacto	113
53. Coeficientes de variables para <i>logisticregression</i> usando punto exacto	113
54. Coeficientes de variables para <i>supportvectormachine</i> usando punto exacto	114
55. Clases de susceptibilidad para evaluación utilizando intervalos iguales	114
56. <i>Landslide density</i> usando media <i>validation decision tree</i>	115
57. <i>Landslide density</i> usando media <i>validation random forest</i>	115
58. <i>Landslide density</i> usando media <i>validation ada boost</i>	115
59. <i>Landslide density</i> usando media <i>validation xgboost</i>	115
60. <i>Landslide density</i> usando media <i>validation logistic regression</i>	116
61. <i>Landslide density</i> usando media <i>validation support vector machine</i>	116
62. <i>Landslide density</i> usando punto exacto <i>validation decision tree</i>	116
63. <i>Landslide density</i> usando punto exacto <i>validation random forest</i>	116
64. <i>Landslide density</i> usando punto exacto <i>validation ada boost</i>	117
65. <i>Landslide density</i> usando punto exacto <i>validation xgboost</i>	117
66. <i>Landslide density</i> usando punto exacto <i>validation logistic regression</i>	117
67. <i>Landslide density</i> usando punto exacto <i>validation support vector machine</i>	117
68. <i>Landslide r index</i> usando media <i>validation decision tree</i>	118
69. <i>Landslide r index</i> usando media <i>validation random forest</i>	118

70. <i>Landslide r index usando media validation ada boost</i>	118
71. <i>Landslide r index usando media validation xgboost</i>	118
72. <i>Landslide r index usando media validation logistic regression</i>	119
73. <i>Landslide r index usando media validation support vector machine</i>	119
74. <i>Landslide r index usando punto exacto validation decision tree</i>	119
75. <i>Landslide r index usando punto exacto validation random forest</i>	119
76. <i>Landslide r index usando punto exacto validation ada boost</i>	120
77. <i>Landslide r index usando punto exacto validation xgboost</i>	120
78. <i>Landslide r index usando punto exacto validation logistic regression</i>	120
79. <i>Landslide r index usando punto exacto validation support vector machine</i>	120
80. Resultados generales usando media	124
81. Resultados generales usando punto exacto	124
82. Resultados de los modelos evaluados por la métrica miou	129

Los derrumbes son un desastre natural que afecta a Guatemala de manera constante y mayormente en temporadas lluviosas. Los mapas de susceptibilidad de deslizamientos son una herramienta crítica para la prevención y mitigación de deslizamientos, así como para la gestión y planificación en el futuro, especialmente para construcciones. Los modelos de *machine learning* son una alternativa al problema de deslizamientos no explorada en Guatemala y este proyecto, presenta un primer acercamiento utilizando modelos de *machine learning* como *Decision tree*, *Random forest*, *XGBoost*, *Logistic regression* y *Support vector machine*. Para ello, se generó un inventario de deslizamientos con 3978 ubicaciones con más de 18 factores predisponentes, entre los cuales figuran los más importantes como la elevación, la pendiente, la clasificación taxonómica del suelo, el tipo del suelo y sus propiedades mecánicas, el tipo de roca, el período de la roca y como factor disparador la precipitación acumulada en un punto de los últimos 7 días. A partir de ello, se pudo entrenar y evaluar los resultados de cada modelo utilizando métricas como *accuracy*, *roc auc*, *landslide density* y *landslide r-index*. Se logró aprobar 4 de los 5 modelos generados, obteniendo valores de *roc auc* superiores a 0.85 y *accuracy* mayores a 0.82, adquiriendo los mejores resultados en el modelo de *XGBoost* con 0.89 de *accuracy* y 0.95 de *roc auc*. Con los modelos, se logra generar mapas de susceptibilidad de deslizamientos en Guatemala a nivel de detalle de 1km. Por último, estos mapas pueden ponerse a disposición de las personas indicadas, como planificadores de proyecto, personal gubernamental e instituciones de prevención a desastres, a través de una plataforma de fácil uso y actualización diaria que permite planear y mejorar el manejo de desastres en Guatemala.

Landslides are a natural disaster that affects Guatemala constantly and mostly in rainy seasons. Landslide susceptibility maps are a critical tool for landslide prevention and mitigation, as well as future planning and management, especially for construction. Machine learning models are an alternative to the landslide problem not explored in Guatemala and this project presents a first approach using machine learning models like Decision Tree, Random Forest, XGBoost, Logistic Regression and Support Vector Machine. For this, an inventory of landslides was generated with 3,978 locations with more than 18 predisposing factors, among which the most important include elevation, slope, taxonomic classification of the soil, the type of soils and its mechanical properties, the type of rock, the period of the rock and as a trigger factor the precipitation accumulated at a point in the last 7 days. From this, it was possible to train and evaluate the results of each model using metrics such as accuracy, roc auc, landslide density and landslide r-index. It was possible to approve 4 of the 5 generated models, obtaining values of auc greater than 0.85 and accuracies greater than 0.82, acquiring the best results in the XGBoost model with 0.89 accuracy and 0.95 roc auc. With the models, it is possible to generate landslide susceptibility maps in Guatemala at 1 km level of detail. Lastly, these maps can be made available to the right people, such as project planners, government personnel and disaster prevention institutions, through an easy-to-use platform that is updated daily to plan and improve disaster management in Guatemala.

Guatemala es muy propenso a desastres naturales, especialmente a derrumbes que ocurren cada vez por los factores permanentes y accidentales del terreno del país. Los factores permanentes incluyen los factores geomorfológicos, geológicos, litológicos y topográficos, mientras que un factor disparador clave es la precipitación constante en las temporadas lluviosas. Estos derrumbes o también llamados deslizamientos de tierra se producen cuando una masa de terreno se convierte en una zona inestable y se desliza respecto a la zona estable. Específicamente ocurren cuando se alcanza la tensión tangencial máxima en todos los puntos en el terreno. En nuestro país, estos derrumbes ocasionan muchas horas de tráfico por bloqueos en las carreteras y no solo esto, también muchas muertes cuando esto sucede en comunidades de personas o ya bien en la misma carretera. El objetivo de este proyecto es encontrar una forma de monitorear zonas susceptibles a deslizamientos, asignando un grado de susceptibilidad de que exista uno de estos para poder evitarlo o ya bien reducir al mínimo sus consecuencias.

Existen muchos libros de ingeniería civil que hablan sobre cómo evaluar el peligro de un deslizamiento de tierra con métodos cualitativos basados en el conocimiento de expertos ya que estos deben conocer la importancia que tienen cada uno de los factores del terreno en la incidencia de estos fenómenos. Existen otros métodos cuantitativos no explorados aún en Guatemala, que se fundamentan en modelos estadísticas y de *machine learning*. A estos últimos, las importancias de las variables son designadas al modelo y, por lo tanto, es más objetivo y otorgan buenos resultados para el mapeo de la susceptibilidad de deslizamientos. Antes de poder considerar estos modelos, se hace necesario crear un inventario de deslizamientos de Guatemala junto con un modelo de datos, el cual contenga información de los factores mencionados anteriormente para que el modelo pueda aprender de forma adecuada. Entre los factores más importantes incluidos se encuentran la elevación, la pendiente, la clasificación taxonómica del suelo, las propiedades mecánicas del suelo, el tipo de roca, el período de la roca y como factor iniciador, la precipitación acumulada en un punto de los últimos 7 días. A partir de ello, se procede a entrenar 5 modelos de *machine learning*: *Decision tree*, *Random forest*, *XGBoost*, *Logistic regression* y *Support vector machine*. Consecuentemente, logrando aprobar 4 de los 5 previos, validando cada uno de los modelos con

métricas como *accuracy*, *roc auc*, *sensitivity* y *specifity*, al igual que métricas estadísticas específicas para el análisis de susceptibilidad como *Landslide Density* y *Landlide Density Relative Index*.

Finalmente, con los modelos de *machine learning* entrenados y los mapas de susceptibilidad generados de Guatemala, se pueden poner a disposición de las personas interesadas a través del diseño de un sistema de monitorización constante que incluye una base de datos que actualiza diariamente la precipitación para poder predecir una susceptibilidad diaria de deslizamientos dependiente de este elemento y los demás factores permanentes, al igual que una interfaz, para visualizar dichos mapas de susceptibilidad. Todo esto permite avances en la zonificación de áreas susceptibilidad a deslizamientos en Guatemala, que tienen gran valor para la prevención y mitigación de deslizamientos, así como para la gestión y planificación futura de estas áreas propensas a estos desastres.

En Guatemala, constantemente se están produciendo derrumbes y deslizamientos, especialmente en la temporada de lluvias. Estos desastres dejan muchas pérdidas humanas y producen numerosos daños en carreteras, propiedades y comunidades. Se hace claro que un sistema de monitoreo de derrumbes para un país como Guatemala podría significar reducir en gran medida los daños que se puedan reducir los daños de este fenómeno, principalmente en cuanto a la reducción de daños las vidas humanas afectadas y los costos económicos de bienes e infraestructura dañada. De igual forma, la utilización de datos de entidades como el Ministerio de Agricultura y Ganadería de Guatemala [MAGA] y el Instituto Nacional de Sismología, Vulcanología, Meteorología e Hidrología de Guatemala [INSIVUMEH], en cuestión de suelos, geología, litología y fisiografía permiten realizar un mejor análisis sobre las áreas que tengan susceptibilidad a un deslizamiento.

¿Por qué los objetivos propuestos para resolver la problemática son viables y aportarán a la solución del problema?

- Los objetivos propuestos pretenden establecer una forma efectiva al monitoreo de posibles derrumbes en zonas de riesgo, de forma que se pueda facilitar las alertas sobre posibles indicios de derrumbes y se puedan establecer estrategias claras a priori para evitar la mayor cantidad de estas calamidades.
- La digitalización y monitoreo constante de derrumbes con imágenes satelitales da facilidad de detección más rápida y optimiza el trabajo del ingeniero civil enfocándose en plantear una propuesta frente a un posible deslizamiento, evitando así pérdidas económicas y humanas.

3.1. Objetivo general

Diseñar e implementar un sistema de monitorización para zonas de riesgo de derrumbes a través de imágenes del área, de modo que permita conocer un grado de susceptibilidad de que estos deslizamientos ocurran y posibilite tomar acciones preventivas.

3.2. Objetivos específicos

- Obtener el ángulo de los taludes a monitorear a través de visión por computador y la ubicación del área analizada por medio de geoposicionamiento.
- Determinar los objetos que se encuentran alrededor de los taludes monitorizados tales como estructuras, carreteras y elementos en la naturaleza.
- Realizar un modelo de datos que permite representar la información útil del área de riesgo.
- Guardar la información histórica en una base de datos de las áreas que han sido monitorizadas.
- Determinar las variables y conceptos más importantes de la ingeniería civil para la detección de la susceptibilidad del posible deslizamiento de un talud.
- Desarrollar un modelo de susceptibilidad de derrumbes que de como resultado un factor de riesgo de ocurrencia.
- Integrar el sistema de monitoreo con los datos de suelo del Insivumeh facilitando así el acceso a esta información y siendo útil para el modelo de susceptibilidad.
- Desarrollar una plataforma que permita monitorizar las áreas por medio de los modelos de datos y de susceptibilidad, informando oportunamente de riesgos de deslizamiento.

4.1. Derrumbes

4.1.1. ¿Qué es un derrumbe?

Los derrumbes son uno de los mayores desastres naturales que existen, entre otros como tormentas, inundaciones, sismos y erupciones volcánicas. Varios geólogos e ingenieros utilizan definiciones particulares sobre lo que es un derrumbe que demuestran la complejidad del estudio relacionado con los mismos. Se ocupará la definición del Servicio Geológico de Estados Unidos [USGS], que explica deslizamiento de forma muy general, como cualquier movimiento descendiente de masa de rocas, escombros o suelo a través de una pendiente por la influencia de la gravedad. Estos movimientos ocurren cuando las fuerzas que actúan en la pendiente exceden la fuerza de los materiales que componen dicha pendiente, causando el deslizamiento. Un deslizamiento puede ser desencadenado por uno o varios factores combinados que se estudiarán más tarde, pero entre ellos se puede mencionar sismos, lluvias, actividad volcánica, erosión, cambios en el nivel del mar o cambios en el suelo por actividad humana. Existen varios términos para hacer referencia a un derrumbe, tales como deslizamiento de tierra, movimientos de masa, fallas de pendiente, entre otros, que se utilizarán indistintamente, aunque más adelante se explicará la clasificación de estos fenómenos.

Estos desastres naturales pueden ocurrir en todas las partes del mundo, con diferentes tipos de terreno y con diferentes situaciones climáticas. Por ello, a pesar de que los sismos y terremotos tengan la mayor cantidad de muertos como desastre natural, ya que suceden a una escala regional, los deslizamientos implican una de los mayores riesgos delimitados a escalas más pequeñas y con mayor frecuencia de ocurrencia (inclusive siendo generados por terremotos haciendo el desastre aún más grande) [1]. Cada año estos implican millones de pérdidas monetarias, así como tristemente la muerte de miles de personas por las condiciones en las que suceden.

Como se mencionó anteriormente, existen diferentes tipos de deslizamiento que deter-

minarán la posible velocidad del movimiento, el volumen de masa que será posiblemente desplazado, así como los probables efectos y mitigaciones apropiadas según el caso. Se pueden clasificar por la forma en que se desplaza el deslizamiento y el tipo de material en el deslizamiento. El material puede ser roca o tierra o una combinación de estas. La tierra puede ser arena fina o escombros si los fragmentos fuesen de mayor tamaño. Los tipos de movimientos es la forma en que se desplaza el material, que puede ser caída, derribo, deslizamiento, expansión o corriente, siendo los más importantes [2]. Con esto, se puede dividir los deslizamientos en dichas categorías ahondadas en la siguiente sección. [1]

Figura 1: Ejemplo de derrumbe de tierra en Guatemala en libramiento de Chimaltenango



Fuente: [3]

4.1.2. Categorías de tipos de deslizamiento

4.1.2.1. Caída

Las caídas se producen en pendientes muy pronunciadas y el movimiento es a gran velocidad, perdiendo a menudo el contacto con el suelo. Su velocidad se basa en partir de un desprendimiento repentino de tierra o roca. El tamaño y el volumen pueden variar desde pequeñas rocas hasta grandes trozos de tierra o roca que pueden romperse con el impacto o continuar a lo largo de las pendientes. Sus causas pueden ser el debilitamiento de la pendiente o el resultado de excavaciones y obras viales o ya bien terremotos, temblores y vibraciones. El esquema de caída de roca puede verse en la Figura [3] y un ejemplo reciente en Guatemala en la Figura [1]

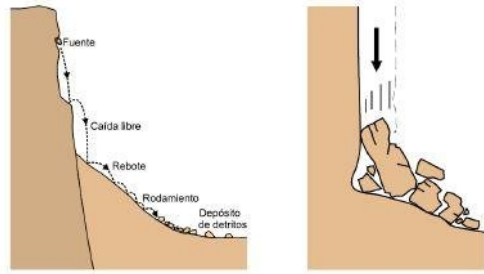
Como resultado, las caídas pueden ser fatales cuando las casas están cerca o debajo, existen carreteras potencialmente causando muertes de personas, además de causar grandes daños a las propiedades y carreteras subyacentes, dañar la infraestructura u obstruir los pasos de vehículos y personas. En un enfoque de remediación y mitigación, se puede realizar cubiertas de protección de caminos, barreras rocosas, muros de contención, remoción de rocas y advertencias al público en general sobre áreas peligrosas. De forma de previsibilidad, se pueden determinar los perímetros de zonas de caídas de roca y cartografía de las zonas de caída de roca peligrosas.

Figura 2: Ejemplo de caída de rocas en Guatemala en carretera de Chiquimula



Fuente: [4]

Figura 3: Esquema de caída de roca

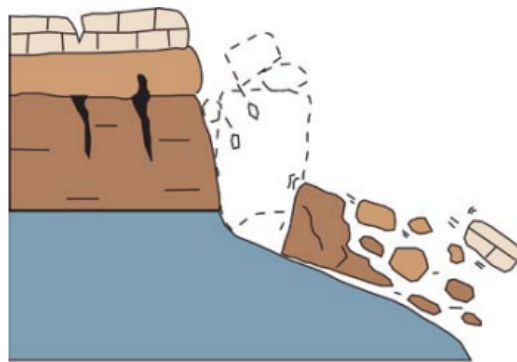


Fuente: [5]

4.1.2.2. Derribo

Un derribo se da por la rotación hacia adelante de una masa que se desprende alrededor de un punto por debajo del centro de gravedad [2]. Estos pueden acarrear roca, escombros o tierra. Su velocidad puede variar bastante, siendo muy lento o muy veloz. Estos se desencadenan impulsados por la gravedad ejercida sobre el material en la pendiente o por erosión, vibración y excavación. Sus efectos dependen en gran manera de la velocidad y del lugar en que sucede el derribo. Su mitigación puede realizarse por refuerzo de estabilización en las zonas y tomar en cuenta la filtración como factor de inestabilidad. El esquema de un derribo puede verse en la Figura [4]

Figura 4: Esquema de un derribo



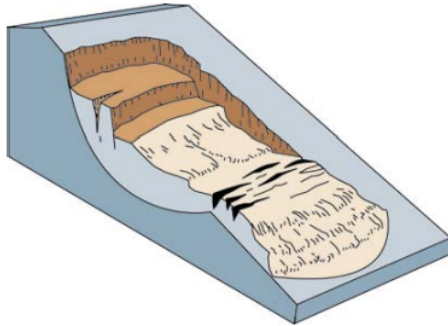
Fuente: [2]

4.1.2.3. Deslizamiento por derrumbe rotacional

Los deslizamientos de tierra ocurren a velocidades que van desde milímetros por mes o año hasta metros por segundo. El daño a ellos se concentra en la superficie de rotura que se curva hacia arriba. Estos están relacionados con la pendiente y la relación con la rotura del suelo. Se abren por fuertes lluvias que provocan la saturación y aumentan la cantidad de agua en la masa o ya bien por la ocurrencia de un sismo [2]. Se observa el esquema de un derrumbe rotacional en la Figura 5 y un ejemplo en Guatemala en la Figura 6

Los impactos dependen en gran medida del tamaño y la extensión del deslizamiento de tierra que causa daños a la infraestructura y las carreteras subyacentes. Las pérdidas mortales dependen de la velocidad a la que sucede el deslizamiento. Como mitigación, se pueden implementar controles para detectar pequeños movimientos, y mediante el cálculo de taludes y pendientes y su gestión se puede reducir significativamente el riesgo, así como el uso de muros de contención. En cuanto a la previsibilidad, es importante reconocer que los registros históricos tienen un alto índice de reincidencia.

Figura 5: Esquema de un deslizamiento rotacional



Fuente: [2]

Figura 6: Ejemplo de un derrumbe rotacional

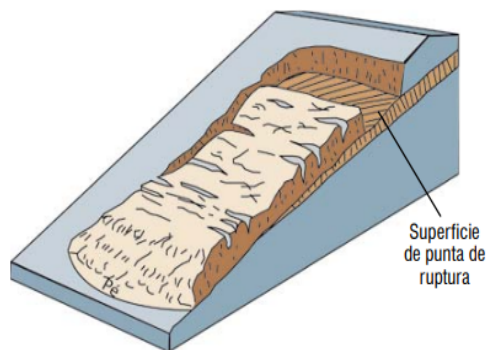


Fuente: [6]

4.1.2.4. Deslizamiento de traslación

Los deslizamientos de traslación suceden cuando la masa se mueva hacia afuera. La distancia que abarca depende de la inclinación y tienden a restablecer el equilibrio de los deslizamientos a diferencia de los de rotación. El tamaño varía desde pequeños tamaños hasta kilómetros. La velocidad también puede ser muy lenta o extremadamente rápida dependiendo de las circunstancias. Son desencadenados por lluvias intensas y el aumento del agua en la masa, al igual que por terremotos. Los efectos y el daño infligido dependen de la velocidad, al igual que los deslizamientos rotacionales. La mitigación radica en drenajes adecuados, muros de contención y medidas correctivas de nivelación. Al igual que con los deslizamientos rotacionales, el riesgo de recurrencia es muy alto. El esquema de un deslizamiento de traslación puede observarse en la Figura 7 y un ejemplo de este tipo en Baja Verapaz, Guatemala en la Figura 8

Figura 7: Esquema de un deslizamiento de traslación



Fuente: [2]

Figura 8: Ejemplo de un derrumbe de traslación en Baja Verapaz, Guatemala

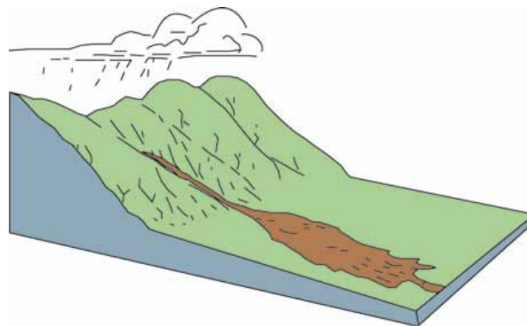


Fuente: [7]

4.1.2.5. Corriente o flujo

Las corrientes o también llamadas flujo de lodo, al igual que los deslizamientos, ocurren a todo tipo de velocidades, desde milímetros por mes hasta metros por segundo. A diferencia de los deslizamientos, estos involucran una fuerte deformación dentro de toda la masa en movimiento. La masa es un líquido viscoso como podría ser el lodo. Son frecuentes en barrancos y se intensifican por el cambio en la tierra por tala o incendios forestales [2]. La velocidad depende la consistencia y la pendiente. Se disparan por flujos intensos de aguas por las lluvias y sus efectos son altos debido a que tienden a velocidades altas por el tipo de material que transportan. Estos son muy difíciles de mitigar, por lo que es importante identificar estas áreas en barrancos porque no se puede construir sobre las mismas. Como previsión, se debe tener mapas de estas áreas de peligro.

Figura 9: Esquema de una corriente o flujo



Fuente: [2]

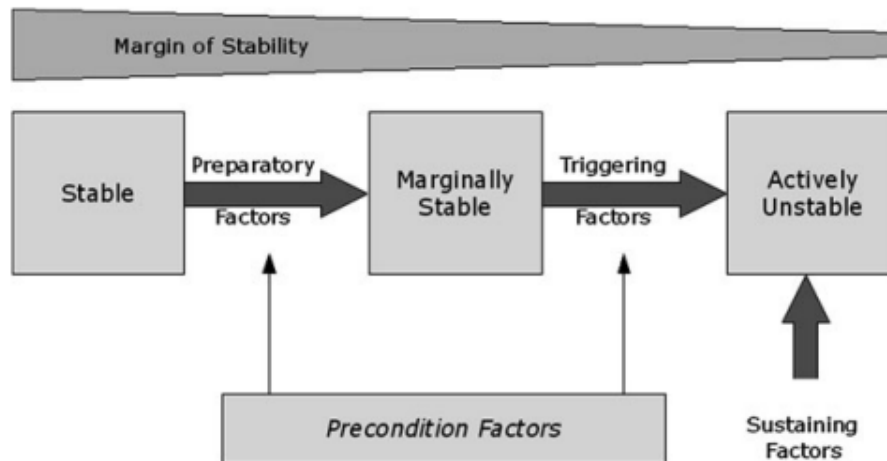
4.1.3. Principios de estabilidad de taludes

La estabilidad de un talud se puede definir como la "propensión a un pendiente para sufrir procesos de deslizamiento de tierra morfológica y estructuralmente disruptivos" [8]. La estabilidad de la pendiente de un talud puede expresarse en términos de la relación entre las fuerzas que impulsan los materiales hacia la pendiente y las fuerzas que resisten la conducción de estrés. Cuando las primeras, exceden a las segundas, es cuando los materiales tienden a moverse y bajar por la pendiente [9]. El factor de seguridad se expresa de la siguiente forma:

$$FS = \frac{ss}{i} \quad (1)$$

donde ss es la resistencia total al corte a lo largo de un plano de corte específico e i siendo el total de cantidad de esfuerzo de estrés desarrollado a lo largo de este plano. Los factores que decrecen esta resistencia son tales como la desintegración y descomposición de rocas por el clima y los procesos hidrológicos; la saturación de materiales, los cambios en la estructura de los materiales. Por otra parte, los factores que incrementan el estrés son la remoción de soportes laterales por ríos o previos deslizamientos; la remoción de los soportes debajo del talud, mayor carga de vegetación y agua; y estrés por vibraciones como sismos. En teoría, el talud es estable mientras el factor de seguridad sea mayor a 1, y si el valor es menor o igual a 1 es los factores preparatorios, más los factores disparadores comienzan la inestabilidad del mismo produciendo un deslizamiento. En la Figura 10, se puede observar el margen de estabilidad de talud al someterse a estos factores.

Figura 10: Estados de estabilidad de un talud



Fuente: [8]

4.1.4. Causas de deslizamientos

La causa de los deslizamientos está ligada a la inestabilidad de los taludes, que como se ha visto anteriormente es cuando el factor de seguridad se vuelve bajo y vulnerable. Se debe distinguir entre las causas y los factores disparadores, siendo las causas los factores

que hacen que un talud se vuelva inestable, mientras que los factores disparadores son los eventos que finalmente iniciaron el movimiento (ver Figura 10) [9]. Por ahora se puede afirmar, que las causas pueden ser naturales o de actividad humana (antropológicas). Las causas naturales incluyen la inclinación de la pendiente, la morfología del terreno, el tipo de suelo y la geología. Los disparadores más importantes son uno o varios tales como el agua, la actividad sísmica y la actividad volcánica, haciendo que Guatemala al tener frecuencia de los 3 disparadores sea una región muy propensa a deslizamientos [2].

4.1.5. Factores causantes de deslizamientos de tierra

La gravedad es el agente principal de los deslizamientos de tierra, siendo este fundamental en la ocurrencia de este fenómeno, pero existen otros agentes importantes que influyen y estos se pueden dividir en cuatro grandes grupos, las cuales son las causas que son las causas de índole geológica, las causas por procesos físicos, las causas morfológicas y las causas de origen antropológico [10].

Los deslizamientos de tierra ocurren por todos los factores que se describen en el Cuadro 1. Sin embargo, algunos de estos elementos son más determinantes que otros en la ocurrencia de este. Como, por ejemplo, los materiales del terreno y la inclinación del mismo [10].

Cuadro 1: Tipos de factores causantes de deslizamientos

Causas de índole geológica	Causas por procesos físicos
Materiales débiles	Precipitación intensa
Materiales sensibles	Actividad sísmica
Materiales en fisuras	Erupciones volcánicas
Discontinuidades orientadas adversamente (inclinación)	Expansión e hidratación de terreno
Discontinuidades estructurales (fallas y discordancias)	
Permeabilidad contrastante	
Contraste de materiales con diferente plasticidad	
Causas morfológicas	Causas de origen antropológico
Levantamientos tectónicos	Excavación de laderas
Erosión fluvial al pie de las laderas	Incremento de peso en las laderas
Erosión subterránea	Disecación de cuerpos de agua (presas)
Remoción de la vegetación	Deforestación
	Irrigación (Canales de agua en plantaciones)
	Actividad minera
	Vibraciones causadas por maquinaria

Fuente: [10]

Estos factores que determinan la ocurrencia de un deslizamiento también se pueden clasificar como permanentes o variables en el terreno. Los ejemplos más importantes de estos factores causantes de deslizamientos se pueden ver en el Cuadro 2.

Cuadro 2: Factores causantes de deslizamientos

Permanentes	VARIABLES
calidad de pendiente	Vibración del suelo
Tipo de terreno	Rápida elevación de nivel de aguas subterráneas
Vegetación del terreno	Humedad en el suelo

Fuente: [10]

Los factores permanentes son características del terreno que permanecen sin cambio o que su cambio no es tan drástico. Estos factores permanentes se pueden reconocer y evaluar luego del deslizamiento, por lo que no es necesario tener estos datos antes de la ocurrencia de este desastre. Mientras que los factores variables son las características del terreno que si cambian con el tiempo y son provocados mayormente por eventos externos. Por lo general, estos factores se tienen que analizar antes de que ocurran estos deslizamientos o justo después de que ocurra, dado que puede variar al poco tiempo de ocurrir el desastre [2].

La mayor parte de las causas de índole geológica y morfológica son de tipo permanente, ya que las geológicas son características del área y las morfológicas reflejan el tipo de mecanismo que les dio origen. Mientras que las causas de procesos físicos y de origen antropológico son características variables en su mayoría, porque los de procesos físicos se refieren a sucesos naturales externos al terreno y las de origen antropológico son los deslizamientos que son provocados por la actividad de los humanos [10].

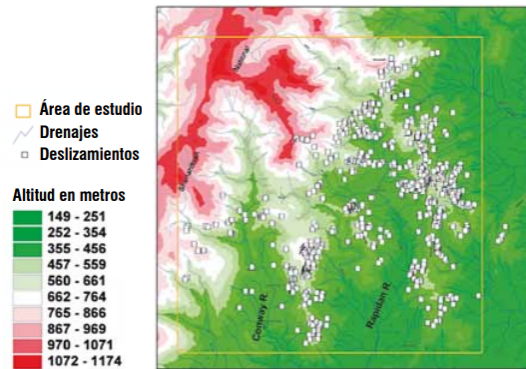
4.2. Importancia de evaluación de peligro de deslizamientos

Los mapas de deslizamientos son herramientas útiles para presentar información sobre el peligro de este desastre natural. Estos son importantes para la planificación, prevención y comunicación con las personas en áreas potencialmente expuestas.

4.2.1. Mapas de inventarios de deslizamientos

Los mapas de inventarios se realizan en áreas que han presentado con anterioridad deslizamientos en ciertas áreas [11]. Estos indican el lugar exacto en el que sucedió uno o múltiples deslizamientos y sirven para diferentes análisis sobre el terreno. Esos análisis varían dependiendo del nivel de detalle que se tenga en el mapa. El nivel de detalle de mapas de inventarios van desde un nivel simple, los cuales muchas veces solo delimitan la zonas donde parece haber deslizamientos, hasta los mapas con un nivel detallado, que suelen tener algunas o toda la siguiente información: estado de actividad, la certidumbre de la identificación, los tipos dominantes de movimiento de las pendientes, el espesor estimado del material de los derrumbes, el tipo de material y las fechas o los períodos de actividad [2].

Figura 11: Ejemplo de un mapa de inventario de deslizamientos



Fuente: [10]

Los mapas de inventario simples sirven para dar una idea general del área en donde ocurrió un deslizamiento, y ayuda a zonificar el terreno de modo que se pueda profundizar en un estudio posterior las zonas más relevantes. Mientras que los mapas de inventario de deslizamientos con un nivel detallado puede proporcionar información más relevante para un análisis más profundo, en el cual se puede comprender los diferentes procesos causantes del deslizamiento, para así identificar posibles áreas propensas a un deslizamiento en un futuro y tomar medidas ya sean correctivas para evitar que suceda un futuro deslizamiento o medidas preventivas e identificar las áreas de terreno estable para el uso de la tierra [2].

4.2.2. Mapas de susceptibilidad

Un mapa de susceptibilidad de deslizamientos nace de un mapa de inventario de deslizamientos, ya que este último es fundamental para su realización. Con el fin de llevar a cabo una relación entre los principales factores que influyeron en el colapso y las distribuciones de deslizamientos en el pasado. Estos no hacen predicciones absolutas, sino que brindan una evaluación cuantitativa o cualitativa de la futura ocurrencia de un deslizamiento en una ubicación [2]. Muchos científicos utilizan el término de mapas de peligros de manera intercambiable con el término de mapa de susceptibilidad; sin embargo, la diferencia es que el segundo no considera la magnitud y la frecuencia para la generación de un modelo. Por último, se tienen los mapas de riesgo que predicen las pérdidas y el daño probable ocasionado por un deslizamiento en una región dada, a través de utilizar los mapas de susceptibilidad, de riesgo, la vulnerabilidad y los elementos expuestos al daño.

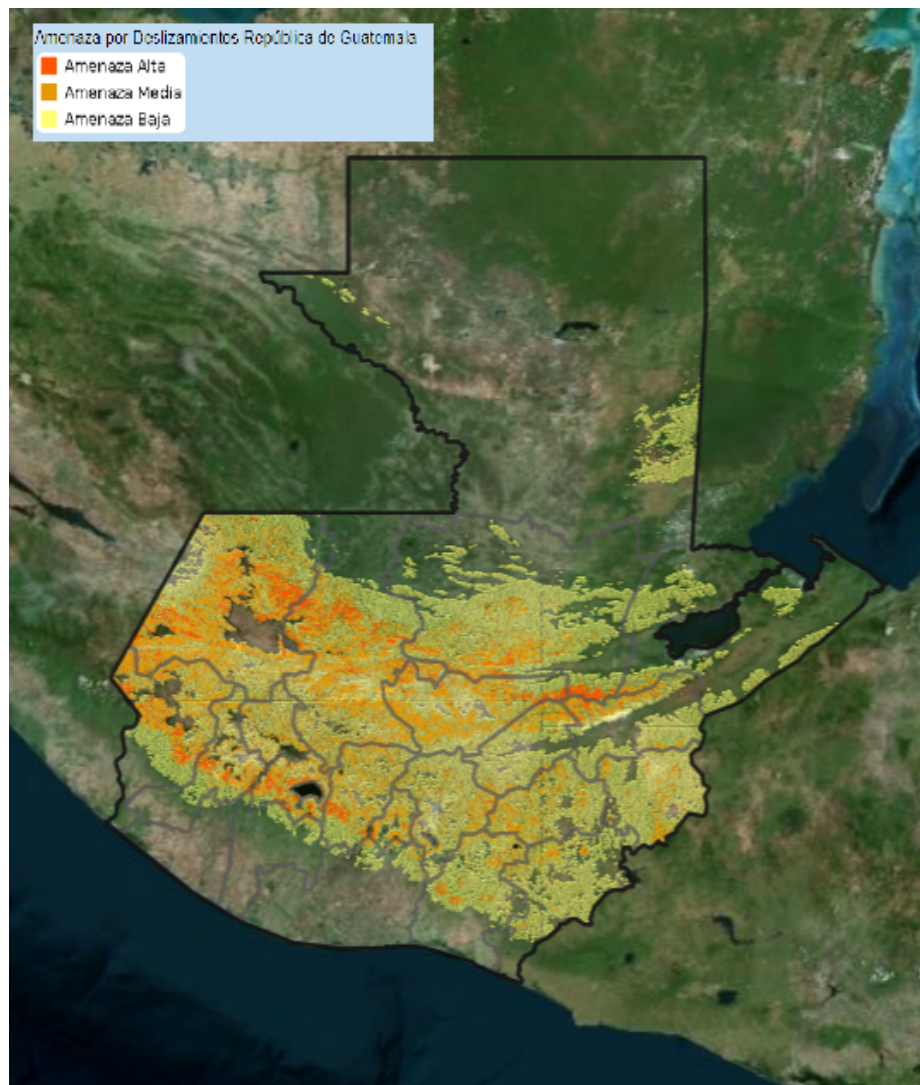
Cabe mencionar, que los mapas de susceptibilidad producidos por el juicio de expertos tienen un problema de subjetividad para la evaluación del peso de las variables de entrada y, por lo tanto, es mejor utilizar análisis estadísticos para encontrar las relaciones de estos parámetros y la *data* del inventario de deslizamientos. Existen varios acercamientos y técnicas para generar mapas de susceptibilidad mencionados en el Cuadro [20], que pueden dividirse en métodos cualitativos, métodos semi-cuantitativos y métodos cuantitativos.

La importancia de los mapas de susceptibilidad radica en reconocer y evaluar los factores que condicionan e inician los deslizamientos para la reducción de peligro futuro y toma de

medidas preventivas y mitigativas. Los principales estudios que se desencadenan son los siguientes [11]:

- Procesos particulares de diseño y medidas para estabilización para mitigar, prevenir o reducir daño.
- Análisis de factores que controlan e inician procesos de deslizamientos en ubicaciones particulares.
- Mapeo de potenciales zonas inestables para delimitar áreas de peligro y tomar medidas preventivas.

Figura 12: Mapa de susceptibilidad de derrumbes en Guatemala



Fuente: [12]

4.2.2.1. Enfoques de mapeo de amenazas y susceptibilidad a deslizamientos

La susceptibilidad a deslizamientos y la zonificación de riesgos, son utilizados para crear zonas basadas en el potencial peligro de que ocurra un deslizamiento. Estos son importantes para una evaluación rápida de la estabilidad de taludes en un área grande y pueden ser útiles para proporcionar información importante sobre una futura ocurrencia de deslizamiento. Para crear estos mapas se han desarrollado varios enfoques que pueden clasificarse en métodos cualitativos, cuantitativos y semi-cuantitativos [11].

Los métodos cualitativos son muy utilizadas en la actualidad, aunque sean enfoques relativamente subjetivos, dado que explican el nivel de condición de deslizamiento en una expresión descriptiva basada en la evaluación de un experto. Dado la naturaleza de estos métodos, para la realización de un método cualitativo se requiere la opinión de especialistas en experiencia del campo [11]. En el Cuadro 3 se mencionan los métodos cualitativos que existen en la actualidad, así como su utilidad y las limitaciones que poseen.

Los métodos cuantitativos se pueden dividir en tres grandes grupos, los cuales son: métodos de *data mining*, métodos basados en la física y métodos estadísticos. Los métodos estadísticos están clasificados en dos grupos, por tener bivariantes y por tener multivariantes [11]. En el Cuadro 4 se mencionan los métodos cualitativos que existen en la actualidad, así como su utilidad y las limitaciones que poseen.

Los métodos semi-cuantitativos son una combinación entre los métodos cuantitativos y los métodos cualitativos. Estos pueden evaluar los efectos de los factores que gobiernan los deslizamientos dado un resultado cuantitativo y con estos resultados realizar análisis cualitativos del terreno y su relación con estos resultados, dado que estos utilizan los dos métodos. Ejemplos de estos métodos se encuentran el proceso de jerarquía analítica, combinación lineal ponderada y conocimiento experto/heurístico. Aunque se introducen algunos conceptos estadísticos en estos métodos, estos aún dependen de la experiencia del experto y del conocimiento profesional, por lo que queda ligado a cierta subjetividad [11].

Cuadro 3: Métodos cualitativos

Método	Descripción	Usos	Limitaciones
Heurístico	Se basa en la opinión y describe un área afectada por un deslizamiento.	Se utiliza para clasificar los mapas de riesgo y susceptibilidad de deslizamientos de tierra.	Depende mucho de la opinión, por lo que tiende a ser subjetivo.
Inventario de deslizamientos	Registra todos los eventos históricos de deslizamientos en un área específica. Y de esta manera se guarda la posición, dimensión, fecha de ocurrencia, entre otros.	Es utilizado como base para las evaluaciones de susceptibilidad, peligro y riesgo de deslizamientos de tierra.	No proporciona la relación espacial entre el deslizamiento de tierra y los conjuntos de factores de deslizamiento de tierra.
Evaluación de factores de peligro de deslizamientos (LHEF)	Evalúa un terreno tomando en cuenta únicamente los factores causantes permanentes.	Se utiliza para la susceptibilidad a los deslizamientos y la zonificación/mapeo de peligros.	No toma en cuenta los factores causantes variables o factores disparadores.
Evaluación de parámetros de estabilidad de taludes (SSEP)	Es una técnica de zonificación del riesgo de deslizamientos que evalúa tanto los factores permanentes, como factores externos o variables.	Se utiliza para preparar mapas de susceptibilidad a los deslizamientos de tierra	Es subjetivo para la asignación de pesos o importancias de los factores de un deslizamiento.

Fuente: [11](#)

Cuadro 4: Métodos cuantitativos

Método	Descripción	Usos	Limitaciones
Análisis estadístico de bivariantes	Es un análisis estadístico que clasifica los deslizamientos en subclases con base en características del talud o terreno para encontrar la relación entre estas subclases y la ocurrencia de un deslizamiento.	Se utiliza para crear mapas de distribución de deslizamientos dado una subclase determinante y también para generar mapas de índices de susceptibilidad de deslizamientos.	n No puede distinguir qué factor es más influyente y no influyente y además no puede predecir cuándo ocurrirá este deslizamiento.
Análisis estadístico de multivariantes	Comúnmente es una regresión logística, un análisis discriminante o un análisis de conglomerados. Obtiene el grado de susceptibilidad de deslizamientos relacionando todas las características del área analizada.	Es utilizado para generar mapas de índices de susceptibilidad de deslizamientos.	No puede predecir cuándo ocurrirá este deslizamiento.
<i>Data mining</i>	Pueden ser modelos de <i>Random forest</i> , <i>boosted regression trees</i> , <i>classification and regression trees</i> , <i>Naïve Bayes</i> , <i>Support vector machines</i> , <i>logistic model trees</i> o <i>artificial neural networks</i> .	Se utiliza para generar mapas de índices de susceptibilidad de deslizamientos.	No puede predecir cuándo ocurrirá este deslizamiento.
Enfoque basado en la física	Incluye límites de equilibrio y modelos numéricos de elementos finitos y se utilizan para calcular el valor cuantitativo de los materiales de pendiente inherentes del factor de seguridad sobre un área en específico.	Se puede aplicar para el análisis de estabilidad de taludes de suelos y rocas.	Es subjetivo para la asignación de pesos o importancias de los factores de un deslizamiento.

Fuente: [11](#)

4.2.2.2. Validación de modelos de mapas de susceptibilidad

El modelo se puede validar aplicándolo varias técnicas de validación como curva de predicción, curva de *rate* de éxito, simple superposición, un gráfico de barras de comparación de porcentaje de deslizamiento de tierra, error relativo, relativo, *receiver operating Characteristics (ROC)* y densidad de deslizamientos [11]. En el Cuadro 5 se puede ver más detalle de dichas métricas.

Cuadro 5: Métricas de evaluación de modelos de susceptibilidad

Métrica	Descripción
<i>Success and Predictive rate Curve</i>	<i>SRC and PRC</i> pueden ser graficadas usando el <i>training</i> de deslizamientos versus la susceptibilidad de deslizamientos utilizando el porcentaje cumulativo del área. El <i>success rate</i> explica que tan bueno es el modelo para clasificar el área de estudio y entrenamiento. El <i>predict rate</i> explica que tan bueno es el modelo para predecir los factores condicionantes del modelo.
<i>Landslide Density</i>	<i>Landslide density</i> es calculado por medio la ecuación $LD = \frac{\text{percentofobservedlandslide}}{\text{percentofpredictedlandslide}}$. Entre mayor sea <i>LD</i> en regiones muy altas de susceptibilidad se confirma que el modelo es acertado y confiable.
<i>Relative landslide density index</i>	<i>Landslide R index</i> es calculado por medio la ecuación $R\text{-index} = \frac{\frac{n_i}{N_i}}{\sum \frac{n_i}{N_i} * 100}$ donde n_i es el número de deslizamientos en una clase de susceptibilidad, mientras N_i es el número de píxeles en esa clase. Mientras el <i>R-index</i> aumente conforme cada clase de susceptibilidad yendo de <i>low</i> a <i>high</i> confirma que el modelo es aceptable y confiable.
<i>Relative Error</i>	El error relativo es una medición que si se encuentra por debajo de 0.5, el modelo es bueno y por encima, no es aceptable. El error relativo puede medirse por la ecuación, $RE = \frac{\sum TNLS - \sum NLS}{\sum TNLS}$ donde <i>TNLS</i> es el total de deslizamientos en una región y <i>NLS</i> es el número de deslizamientos en muy muy alta susceptibilidad.
<i>Receiver operating Characteristics</i>	La curva <i>ROC</i> es usada para evaluar el <i>performance</i> de un mapa de susceptibilidad. La curva <i>ROC</i> es una representación de los <i>TPR</i> en axis Y y los <i>FPR</i> en el axis X. El área bajo la curva <i>AUC</i> es el valor de diagnóstico, que si se encuentra entre 0.9 y 1 es un indicador de excelente <i>performance</i> , entre 0.8 y 0.9 de un muy buen <i>performance</i> , y de 0.7 a 0.8 como <i>performance</i> bueno. Entre 0.6 y 0.7 es modelo normal y todo lo debajo de 0.5 es un modelo pobre e inútil.

Fuente: [11]

4.3. Sistema de información geográfica (*GIS*)

Es un conjunto de herramientas que integra componentes que permiten la gestión, almacenamiento, manipulación, análisis y modelización de datos espaciales georeferenciados [13].

4.3.1. Funciones principales

4.3.1.1. Almacenamiento

Un *GIS* permite codificar y almacenar los diferentes fenómenos que aparecen en la superficie terrestre. Esto lo hace a través de modelos de datos adecuados, los cuales implican modelizar la realidad y codificar de forma cuantitativa este modelo [13].

4.3.1.2. Visualización

Los *GIS* manejan datos espaciales, los cuales se pueden presentar en 3 dimensiones (x, y, z) o en 4 dimensiones (3 espaciales y el tiempo) o más comúnmente en proyección sobre el espacio bidimensional definido mediante coordenadas cartesianas [13].

4.3.1.3. Consultas

La obtención de respuestas a una serie de consultas sobre los datos y su distribución en el espacio. Una consulta a una base de datos implica:

- Seleccionar el subconjunto de datos que el usuario necesita en función de un conjunto de criterios previamente definidos. Un ejemplo podría ser todos los puntos X Y con una altura de más de 100 m del nivel del mar.
- Presentarlo al usuario de forma útil, bien sea tablas (con listados de los puntos XY), gráficos o mapas en los que los puntos X Y aparezcan de un determinado color. Las tablas dan una información más exacta, pero los mapas presentan aportan información espacial.

[13]

4.3.1.4. Análisis

Una funcionalidad más sofisticada es el uso de herramientas de análisis espacial y álgebra de mapas para el desarrollo y verificación de hipótesis acerca de la distribución espacial de las variables y objetos. En algunos casos resulta necesaria la utilización de programas de análisis estadístico externos a los programas de SIG, debe buscarse entonces la mayor integración posible entre ambos tipos de programas en cuanto a tipos de datos manejados

y compatibilidad de formatos, un ejemplo de estos casos podría ser: ¿Disminuye la temperatura con la altitud?. En otros casos se tratará de implementar modelos ya formulados y fundamentados, como por ejemplo: ¿Cuál es el tamaño mínimo de un área de bosque para mantener una población viable de lince ibérico? [13].

4.3.1.5. Toma de decisiones

Los *GIS* más completos, se utilizan para toma de decisiones mediante el uso de instrucciones y especificaciones del análisis espacial y álgebra de mapas. Por lo que esta funcionalidad puede ayudar al usuario a tomar una decisión sobre un trabajo o proyecto en específico, como por ejemplo le puede ayudar al usuario a decidir dónde construir una estructura pesada en un terreno montañoso [13].

4.3.1.6. Modelación

Las aplicaciones de los *GIS* están ligadas a la anexión de modelos matemáticos y procesos naturales, dinámico y espacialmente distribuidos, es decir, que se pueden integrar objetos científicos o de planificación y ordenación, por ejemplo qué consecuencias ambientales puede tener un estancamiento de aguas abajo de su ubicación o bien, cuál será el impacto sobre el medio de una actividad. Los *GIS* deben ser integrados con un modelo dinámico y llevarse a cabo de varias maneras [13]:

- Implementación del modelo en un programa externo que importa y exporta los formatos de fichero del *GIS*, siendo los programas totalmente independientes.
- Implementación de programas integrados compartiendo el mismo formato de ficheros con la posibilidad de ejecutarse simultáneamente.
- Modelo incorporado como un módulo del *GIS*.

4.3.2. Tipos de archivos manejados en un *GIS*

4.3.2.1. GeoTiff (.tiff)

GeoTIFF es un estándar creado por el Dr. Niles Ritter en el proyecto de la NASA Jet Propulsion Laboratory. Este contiene información georeferenciada encajada en un archivo de imagen de formato TIFF. La información adicional a la imagen incluye el tipo de proyección, sistema de coordenadas y todo lo necesario para que la imagen pueda ser automáticamente posicionada en un sistema de referencia espacial [14].

4.3.2.2. Modelo digital de elevación (*DEM*)

Un *DEM* es una representación de una superficie continua, que en el mayor de los casos hace referencia a la superficie de un terreno, y su valor en cada punto (x,y) representa la

altitud de este terreno con respecto al nivel del mar. La precisión de estos datos se determina principalmente por la exactitud y la resolución horizontal o grado de detalle digital de representación en formato digital. Estos valores dependen del método que se utiliza para generar este modelo [15].

4.3.2.3. Ortomosaico

Una imagen de ortomosaico es una imagen aérea de alta resolución tomada por un UAV (Vehículo aéreo no tripulado). Cuando se realiza una ortorectificación a este tipo de imágenes, estas pueden utilizarse para crear un mapa muy detallado, sin distorsiones, y de esta manera mejorar la visibilidad de los detalles. La ortorectificación elimina la perspectiva de cada imagen individual para crear consistencia en todo el mapa, manteniendo el mismo nivel de detalle de la imagen original [15].

4.3.2.4. Shapefile

Un *shapefile* es la forma estándar de almacenar e interactuar con datos espaciales. Debido a que los *shapefiles* no tienen la sobrecarga de procesamiento de una estructura de datos topológicos, tienen ventajas sobre otras fuentes de datos, como una mayor velocidad de dibujo y capacidad de edición para un *GIS*. También suelen requerir menos espacio en disco y son más fáciles de leer y escribir [16].

Un *shapefile* almacena geometría no topológica e información de atributos para las *features* espaciales de un conjunto de datos. Por lo que cada *feature* dentro de un *shapefile* contiene atributos no espaciales que pueden contener cualquier tipo de información, así como también un atributo "geometry" el cual se almacena una forma que comprende un conjunto de coordenadas vectoriales. Este atributo tiene soporte de puntos, líneas y áreas (las cuales son representadas como un circuito cerrado de coordenadas) [16].

4.3.3. Almacenamiento de información en un *GIS*

En los *GIS* de primera generación, en un inicio todos los datos espaciales se almacenaban en archivos planos como los *shapefiles* y se requería un *software GIS* especial para interpretar y manipular estos datos. También se requería de un nivel de procesamiento mayor para mantener, manipular, iterar y filtrar los datos. Para la segunda generación de los *GIS*, estos almacenaban cierta cantidad de *data* en bases de datos relacionales (usualmente la información no espacial), pero aún no poseía una integración completa para la información espacial [17].

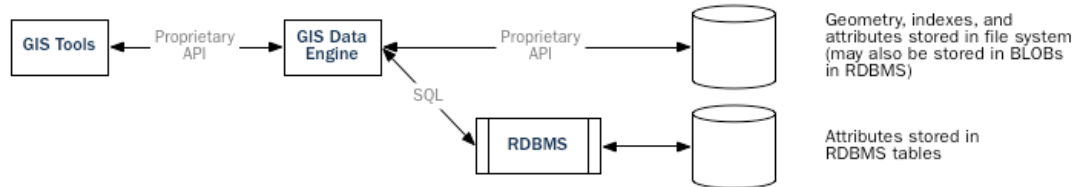
La tercera generación de los *GIS* se creó una integración completa para el almacenamiento de toda la información por medio de bases de datos espaciales. Estas permitieron la integración completa de datos espaciales con bases de datos relacionales tradicionales, por lo que esto hizo que el sistema cambiara de estar centrado en *GIS* a estar centrado en base de datos [17].

Figura 13: Evolución del almacenamiento de datos de *gis*

First-Generation GIS:



Second-Generation GIS:



Third-Generation GIS:



Fuente: [17]

4.3.4. Digital Terrain Analysis (DTA)

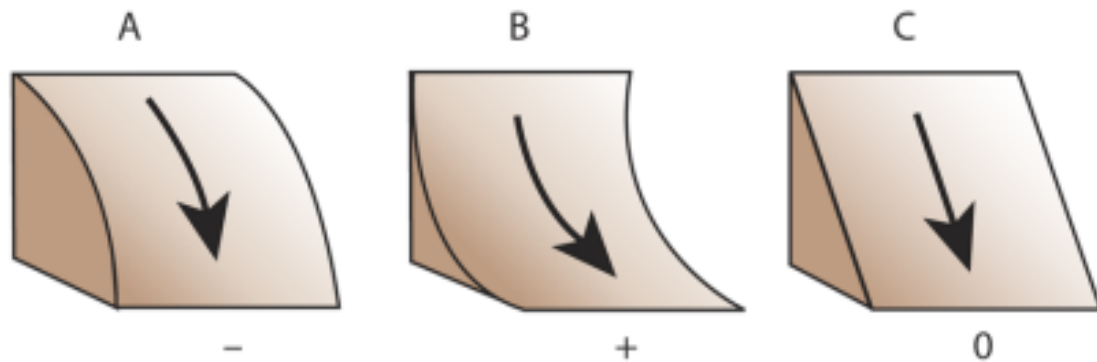
El análisis de los modelos digitales de la topografía de un terreno, como los modelos de elevación digital (*DEM*), son una parte integral de un *GIS*. Estos análisis proveen datos cuantitativos de la morfología de un terreno y pueden ser utilizados para mapear terrenos y ser analizados. Estos datos cuantitativos de la superficie terrestre puede darse en términos de pendiente, aspecto y curvatura de la pendiente y estos brindan información digital del terreno relevante al momento de analizar un terreno en el que existió un deslizamiento [18].

La pendiente mide el grado de cambio de elevación en el terreno. Mientras más grande es la pendiente, la gravedad ejerce una mayor fuerza a todo lo que está sobre el terreno, esto incluye flujo de agua y otros materiales. Por lo que tiene una gran importancia en hidrología y geomorfología, ya que afecta la velocidad en el flujo de la superficie y la subsuperficie, en el potencial de una erosión, en la formación del suelo, entre otros. El aspecto es la orientación de la pendiente. Este es muy útil para la visualización de deslizamientos, especialmente cuando se utiliza un buen color en la escala [18].

Existen dos tipos de curvatura de un terreno: la curvatura de perfil y la curvatura del plano. La curvatura de perfil es afectada y se forma por la aceleración y desaceleración de la pendiente [18]. Como se puede ver en la Figura [14], las curvaturas de perfil negativas (A) indican pendientes convexas hacia arriba, las curvaturas de perfil positivas (B) indican superficies cóncavas hacia arriba y una curvatura de perfil de valor cero indica una pendiente

lineal (C).

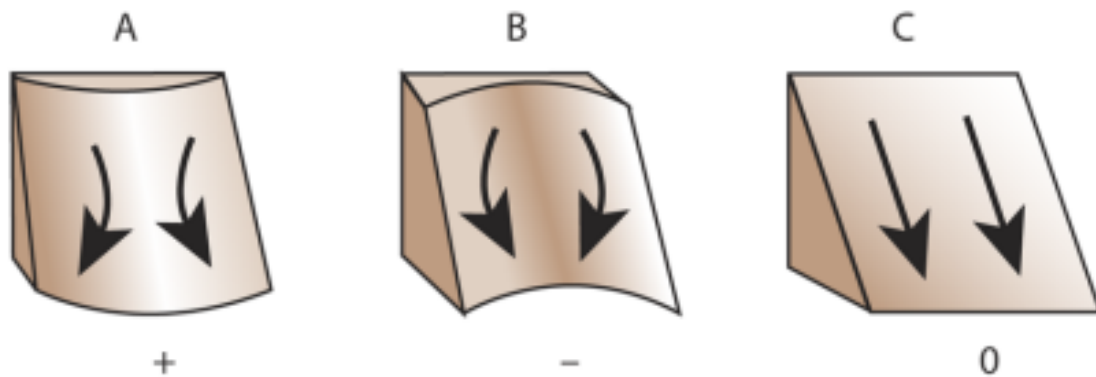
Figura 14: Curvatura de perfil



Fuente: [18]

La curvatura del plano es perpendicular a la pendiente de la superficie y es afectada por la convergencia y divergencia de la pendiente [18]. Como se puede ver en la Figura [15], las curvaturas del plano negativas (A) indican pendientes lateralmente convexas, las curvaturas del plano positivas (B) indican superficies lateralmente cóncavas y una curvatura del plano de valor cero indica una pendiente lineal (C).

Figura 15: Curvatura del plano



Fuente: [18]

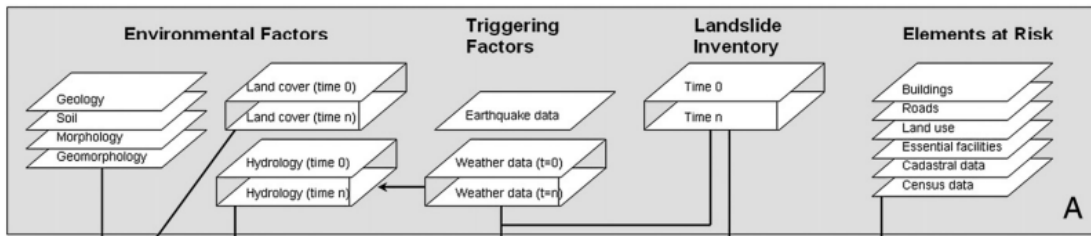
4.4. Factores espaciales para la evaluación de derrumbes

Desde hace mucho tiempo atrás se han realizado grandes esfuerzos para desarrollar modelos de evaluación sobre los deslizamientos. Los ingenieros y geólogos han recurrido a una

serie de herramientas que van desde simples ecuaciones hasta complejos modelos numéricos tridimensionales. Estos enfoques tienen como objetivo evaluar la estabilidad estática de los taludes o predecir el resultado de los deslizamientos. Sin embargo, muchos de los modelos no tienen en cuenta todos los factores, como por ejemplo tomar en cuenta que un gran porcentaje de los derrumbes son causados por lluvias intensas y sismos [1]. Por ello, es importante definir los factores más importantes a tomar en cuenta para crear un modelo de susceptibilidad de derrumbes.

Cees van Westen ha realizado varias investigaciones y estudios sobre la *data* espacial más importante a considerar para realizar una evaluación de susceptibilidad, riesgo y vulnerabilidad de deslizamientos [19]. En la Figura 16, se pueden observar estos factores tales como factores ambientales que incluyen geología, suelo, uso de la tierra, morfología y geomorfología; factores disparadores como *data* de terremotos y *data* climática e hidrológica; un inventario de deslizamientos y finalmente, para evaluar el riesgo datos de censo, uso de tierra, infraestructura, caminos, entre otros. Otros estudios, como los de Wubalem, llevan a los mismos factores afirmando que las variables litológicas, geomorfológicas, estructura geológicas, hidrológicas, climatológicas, antropológicas, sísmicas, los parámetros de uso/cobertura de la tierra y los inventarios detallados de deslizamientos de tierra son los más importantes variables de entrada para el mapeo de susceptibilidad a deslizamientos basado en *GIS* [11]. Mas abajo se entrará a más detalle de cada uno de estos.

Figura 16: Factores importantes para la evaluación de susceptibilidad de derrumbes



Fuente: [19]

4.4.1. Factores ambientales

En los factores ambientales, se incluyen los factores de geología, suelo, morfología y geomorfología.

4.4.1.1. Factores geomorfológicos: *DEM: elevation y slope*

Los factores geomorfológicos están encargados de estudiar el relieve terrestre y estos son de suma importancia para obtener datos muy importantes en la base de datos de derrumbes sobre cómo se encontraba el terreno antes del derrumbe y también cómo se encuentra a día de hoy para reconocer lo que ha cambiado. Se mencionan algunos de los más importantes y lo que pueden brindarnos.

DEM se refiere a *Digital Elevation Model* que una importante modelo para los sistemas de información geográfica (*GIS*). Contiene información sobre la elevación de un terreno y se encuentran otros términos que se relacionan como *Digital Terrain Model* (DTM) y *Digital Surface Model* (DSM). Un DSM incluye la elevación de las alturas de "todo" (árboles, edificios, etc.), mientras que el *DEM* únicamente tiene la altura fuera de dichos objetos. Un DTM contiene más información sobre el terreno, por lo que se podría decir que está conformado de varios *DEM* [20]. Un *DEM* busca representar entonces información sobre la elevación de un relieve. Para obtener esta información se requieren de estudios de campo, por lo general a través de procesos de alto costo y de mucho tiempo. Actualmente, se han desarrollado más tecnologías y nuevos métodos que han optimizado la generación de *DEM* que incluyen fotogrametría digital, teledetección por satélite, LiDAR, InSAR, entre otros.

Digital Terrain Analysis (DTA) es el proceso de la extracción y computación de propiedades derivadas del *DEM*. Se agrupan en 4 tipos de las características según su relación con el terreno, siendo estos grupos, parámetros de la superficie, características morfológicas, atributos estadísticos y atributos compuestos. Los que más nos interesan para la evaluación de derrumbes son los parámetros de la superficie que tienen cantidades escales y tiene una definición física. Estos se encuentran en el Cuadro 6. Como se puede ver, los elevación y pendiente son factores importantes que pueden determinar la velocidad de los deslizamientos, al igual que su inestabilidad.

Cuadro 6: Factores geomorfológicos importantes para el análisis de derrumbes

Factor	Medición	Descripción
Elevación	m	Referente a la cantidad de distancia vertical entre el punto de la superficie y la Geoide.
Área	m^2	Referente la cantidad de área proyectada por unidad de terreno.
Pendiente	$^\circ$	Referente al ángulo entre el vector normal y él de la cima de un punto.
Aspecto	$^\circ$	Haciendo referencia al ángulo entre el vector normal proyectado en el horizonte y el norte. La orientación de la pendiente.
Curvatura	m^{-1}	Haciendo referencia al perfil de la curvatura a lo largo de la pendiente y también a la media. La rugosidad del terreno.

Fuente: elaboración propia

Los *DEM* se pueden obtenerse de satélites de alta resolución (mayor a 10m), pero estos por lo general tienen un alto costo de entre 30 y 50 dólares por kilómetro cuadrado, lo cual se vuelve aún más complejo con la necesidad de imágenes multitemporales y por los altos costos, que aún siguen siendo una limitación hoy en día. Una forma sencilla de obtener *DEM* de no tan alta precisión, pero válidos para un estudio de derrumbes, es a través de los catálogos abiertos de Google Earth Engine. Google Earth Engine es una plataforma abierta (de uso gratuito para fines académicos investigativos) de Google que contiene varios catálogos de imágenes satelitales y *datasets* geo espaciales que son utilizados por científicos e investigadores para múltiples propósitos en la detección cambio en la superficie tierra [21]. De igual forma, contiene un API para Python, fácil de utilizar para nuestros propósitos que serán explorados en la metodología.

Los catálogos de Google Earth Engine que nos interesan para la generación de *DEM* son los siguientes con su descripción:

- Sentinel: Sentinel una misión de imágenes multiespectrales de alta resolución y área amplia con una frecuencia de revisión global de 5 días. El instrumento multiespectral S2 (MSI) muestrea 13 bandas espectrales: visible y NIR a 10 m, borde rojo y SWIR a 20 m, y bandas atmosféricas a una resolución espacial de 60 m [21].
- Landsat, un programa conjunto del USGS y la NASA, ha estado monitoreando continuamente la tierra desde 1972. Hoy en día, los satélites Landsat combinan datos cuantitativos y térmicos para obtener imágenes de toda la superficie de la tierra con una resolución de 30 metros una vez cada dos semanas [21].
- NASADEM *Digital Elevation* 30m es un reprocesamiento de datos STRM, con precisión mejorada al incluir datos auxiliares de conjuntos de datos ASTER, GDEM, ICESat GLAS y PRISM [21].

En el Cuadro 7, se pueden observar el resumen de sus características.

Cuadro 7: Resumen de características catálogos disponibles en Google Earth de imágenes satelitales y *DEMs*

Catálogo	Resolución	Fecha inicio	Fecha fin	Bandas importantes
Sentinel 2	10m	2015-06-23	2022-10-17	B2 Blue B3 Green B4 Red B8 NIR QA60 CloudMask
Landsat 8	30m	2013-03-18	2022-10-10	B2 Blue B3 Green B4 Red B5 NIR QAPIXEL QABitmask
Landsat 7	30m	1999-06-29	2022-09-16	B1 Blue B2 Green B3 Red B4 NIR QAPIXEL QABitmask
NASADEM	30m	2000-02-11	2000-02-22	<i>Elevation</i>
GSMaP Operational	30m	2000-02-11	2022-10-16	hourlyPrecipRate

Fuente: elaboración propia

4.4.1.2. Factores geológicos litológicos

Los factores litológicos se encuentran divididos en los indicadores en los mapas elaborados y recopilados por el Ministerio de Agricultura de Guatemala a una escala de 1:250000 [22], que pueden observarse en el Cuadro 8. Estos factores pueden dividirse en el tipo de roca y

el período al que pertenece el tipo de roca. Diferentes litologías pueden ser un factor para mostrar diferentes grados de susceptibilidad a potenciales fallos.

Cuadro 8: Resumen de propiedades litológicas obtenidas del mapa geológico de Guatemala

Factor	Clasificación
Tipo de roca	Rocas Sedimentarias Rocas Ígneas y Metamórficas
Periodo	Paleoceno-eoceno Eoceno Aluviones Cuaternarios Cuaternario Cretácico Cretácico-terciario Cretácico-eoceno Terciario Terciario superior Oligoceno-mioceno Terciario superior Oligoceno-plioceno Paleozoico Carbonífero-pérmico Jurásico-cretácico Pérmico

Fuente: [22]

4.4.1.3. Factores de suelo e hidrogeológicos

Las propiedades de los suelos son un aspecto muy importante ante el grado de susceptibilidad de derrumbes ya que el posible fallo de la pendiente depende la fuerza de la superficie, que a su vez depende de la cohesión y del ángulo interno de fricción del suelo [11]. Es por ello que en la Cuadro [10] se encuentran las taxonomías de suelo de Guatemala y en el Cuadro [11] se pueden encontrar variables importantes como el riesgo de erosión en el suelo, el material de origen del suelo, las texturas superiores y su calidad del suelo, las texturas inferiores y su calidad del suelo y el drenaje interno del suelo. Todas estas variables son tomadas en cuenta ya que la debilitación de la resistencia de la superficie depende de la naturaleza y el origen del suelo, su tipo y espesor de relleno, persistencia, entre otros [11]. El riesgo de erosión y el drenaje interno del suelo, también son factores ampliamente reconocidos por junto a la precipitación afectar la estabilidad del talud [8]

4.4.1.3.1. Clasificación taxonómica del suelo

La clasificación de suelos siempre ha sido un factor clave para el desarrollo de los estudios y la ciencia de los suelos. El U.S. *Soil* Conservation Service, inicio programas para la construcción de un sistema de clasificación de suelos en el año 1951, y a partir de la publicación en 1999 de *Soil* Taxonomy System, es el sistema de clasificación más utilizado hoy en día junto al sistema propuesto por la United Nations Food and Agricultural Organisation (FAO)

[23]. El sistema de clasificación taxonómica del suelo clasifica el suelo en 6 categorías:orden, Suborden, Grupo, Subgrupo, Familia y Serie, siendo la primera la más amplia y la última, la más específica.

En la clasificación taxonómica por orden, se reconocen 12 órdenes, basados en tres criterios: características morfológicas: unidades de superficie (epipedón) y unidades de sub-superficie ; propiedades químicas: saturación; y características ambientales: regímenes de humedad y temperatura. La clasificación de orden casi solamente se rige por las características morfológicas. Estos órdenes pueden ser encontrados en todo el mundo con patrones reconocibles y a continuación se describen los encontrados en Guatemala [23]:

- Alfisoles: Se desarrollan bajo temperaturas de bosques húmedos y subhúmedos de la tierra. Son suelos bien desarrollados con mayor de 35 % de saturación base.
- Andisoles: Son texturas de suelo finas desarrolladas de material volcánico. Tiene una baja base de saturación.
- Entisoles: Estos suelos son pobres, poco profundos, de origen reciente y fueron desarrollados sobre material no consolidado. Tiene baja capacidad de retención de agua y bajo contenido de material orgánico. Estos tienden a formarse en superficies jóvenes como aluviones, coluviones y flujos de lodo. La deforestación y otras actividades humanas potencian la formación de entisoles. Son comúnmente encontrados en suelos inestables y pendientes pronunciadas.
- Inceptisoles: Estos suelos tienen un desarrollo intermedio, con mayor desarrollo que los entisoles. La creación predomina en pendientes grandes y en superficies jóvenes. Sus drenajes pueden ser muy pobres o muy buenos en pendientes pronunciadas. Los climas con bajas temperaturas y bajas precipitación favorecen el desarrollo de estos suelos. Son predominantes en zonas montañosas.
- Mollisoles: Estos son suelos negros orgánicos con una saturación base mayor al 50 %. Estos ocurren en depósitos de varios años y son fértiles a plantaciones.
- Ultisoles: Son similares a los alfisoles, pero tienen un baja saturación base, por lo general, menor al 35 %. Puede formarse tanto en pendientes pronunciadas como un lugares planos.
- Vertisoles: Estos suelos tienen bastantes arcillas expansivas, que les hacen cambiar con el cambio en la humedad. En periodos secos, crean amplias grietas generadas al encogerse.

4.4.1.3.2. Propiedades físicas del suelo

El suelo posee materiales sólidos, líquidos y gaseosos. La variabilidad de estos componentes determina una serie de propiedades que se conocen como propiedades físicas del suelo, las cuales son: color, textura, estructura, porosidad y consistencia [23]. Más adelante se describen cada una de ellas.

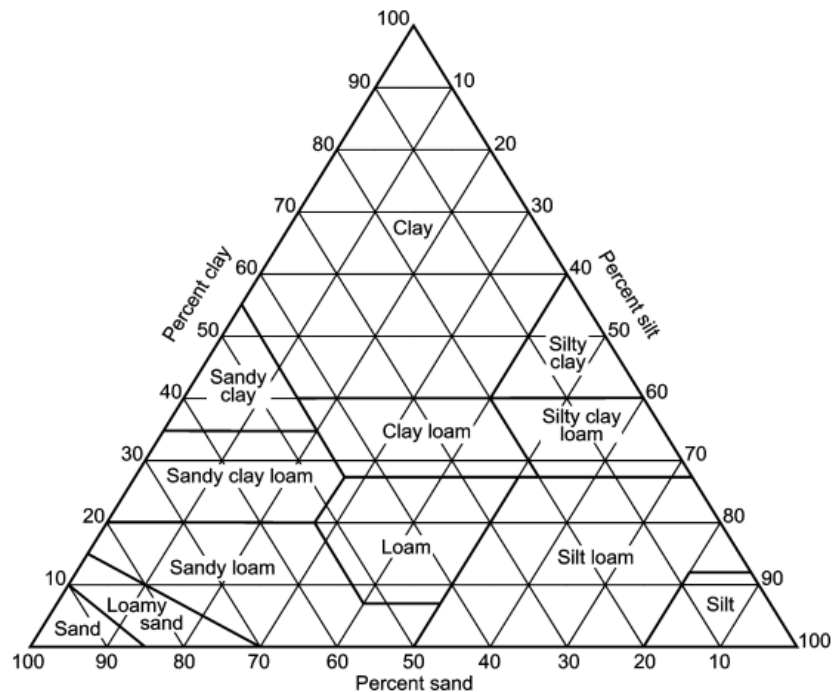
4.4.1.3.2.1. Color

El color es el rasgo más obvio en un perfil de suelo, pero aun así puede ser indicativo de composición del suelo. Los tonos de color rojo y amarillo son generalmente indicativos de enriquecimiento en hierro férrico, mientras que los tonos grises pueden resultar de concentraciones más altas de óxidos de aluminio y silicatos. Los colores negros generalmente son causados por una abundancia de material orgánico [23].

4.4.1.3.2.2. Textura

La textura es el término que se refiere al tamaño y organización de las partículas del suelo. Los suelos pueden clasificarse según su textura en diferentes clases dependiendo de la proporción de arena, arcilla y limo (Fig. 17). Un suelo hipotético formado por partes iguales de estos 3 componentes se denomina franco [23]. Todos los suelos constan de una mezcla de partículas o agrupaciones de partículas de tamaños similares, por lo que también se usa su clasificación con base en los límites de diámetro en milímetros (Cuadro 9). Esta propiedad ayuda al momento de analizar la estabilidad de un talud, ya que al disminuir el tamaño de las partículas del suelo y tener una textura muy fina, disminuye la estabilidad de dicho talud.

Figura 17: Clasificación de textura de suelos



Fuente: [23]

Cuadro 9: Clasificación de las partículas del suelo

Nombre	Tamaño
Arena	mayor a 2.0
Muy gruesa	1.0 - 2.0
Gruesa	0.5 - 1.0
Mediana	0.25 - 0.5
Fina	0.1 - 0.25
Muy fina	0.05 - 0.1
Limo	0.002 - 0.05
Arcilla	menor a 0.002

Fuente: elaboración propia

4.4.1.3.2.3. Estructura

La estructura del suelo se refiere a la disposición de las partículas primarias del suelo para la formación de agregados. Los agregados se separan entre sí a lo largo de las zonas de debilidad, creando vacíos que son importantes para el almacenamiento y drenaje de la humedad. Los agregados esféricos tienen más espacio poroso y mayor permeabilidad. La estructura es importante porque puede modificar o superar parcialmente los aspectos de la textura del suelo. La clasificación que se utiliza para la estructura es: granulares, en bloques, prismáticas y laminar [23].

4.4.1.3.2.4. Porosidad

El tamaño del grano y el volumen de los poros son los factores más importantes que controlan la percolación del agua y la ventilación dentro del suelo. La porosidad es el porcentaje de volumen del espacio poroso con respecto al volumen total. La porosidad también se puede determinar directamente por el volumen de agua contenido en un núcleo de suelo saturado no perturbado de un volumen dado [23].

4.4.1.3.2.5. Consistencia

La consistencia se refiere a las propiedades de cohesión del suelo, como la resistencia mecánica del estrés y la manipulación bajo las condiciones de humedad. La consistencia se expresa como seca, húmeda y mojada. [23].

Los factores de suelo se encuentran divididos en los indicadores en los mapas elaborados y recopilados por el Ministerio de Agricultura de Guatemala a una escala de 1:250000 [22], que pueden observarse en el Cuadro [10] y en el Cuadro [11].

Cuadro 10: Resumen de propiedades taxonómicas de suelos obtenidas del mapa taxonómico de suelos Guatemala

Factor	Clasificación	Porcentaje
Orden taxonómico del suelo	Alfisoles	11.86
	Andisoles	10.05
	Entisoles	20.01
	Inceptisoles	15.00
	Mollisoles	18.23
	Ultisoles	9.74
	Vertisoles	14.8

Fuente: 22

Cuadro 11: Resumen de propiedades de reconocimiento de suelos Simmons obtenidas del mapa de reconocimiento de suelos Guatemala

Factor	Clasificación
Símbolo suelo	Tm
	Cr
	Te
	Tq
	Qt
	Am
	Mj
	Chu
	Tl
	Pi
	Ta
	Ab
	Jl
	Chg
	Ci
	Ca
	Ic
	Ja
	Pz
	Si
	Cme
	Sa
	Tp
	Cm
	Sl
	Qi
	Ce
Ae	
Ay	
Fr	

Factor	Clasificación
	Cc
	Zc
	Qe
	Gt
	Sub
	AGUA
	Sn
	Mg
	Mr
	Cq
	Chv
	So
	Chx
	Tc
	Sx
	Ba
	Bb
	Cu
	Co
	Ye
	Ac
	Po
	Tn
	At
	Pa
	Pt
	Pl
	Ss
	Chr
	Ne
	Qx
	Mq
	Cha
	Cul
	Mo
	Sb
	Sr
	Gp
	Gl
	Sh
	Ju
	Pc
	Qr
	Su
	In
	Chc
	Tz

Factor	Clasificación
	Yx
	Pu
	Cx
	Sz
	Sc
	Ma
	Chl
	Ql
	Ib
	Mp
	Mc
	Chh
	Jo
	Ya
	Pe
	Sp
	Jg
	Chj
	As
	Za
	Gu
	Ji
	Ga
	Zr
	Mi
	Cp
	Ex
	Re
	Ti
	Ix
	Cho
	Mt
	Mz
	Pn
	Sm
	Gc
	Ou
	Chm
	Cs
	Ts
	Bu
	Tt
	My
	Px
	Tx
	Tj
	Ua

Factor	Clasificación
	Qa AM Sui Chq Us Bo Cg Pp Ah Chp Se To Gn Chi Es Xa Tr Cj Chn Cy Suc Cl St Au Oq Ct
Drenaje suelo	Buen Drenaje Drenaje Moderado Drenaje Imperfecto Drenaje Pobre Drenaje muy Pobre Agua
Material origen	Caliza Ceniza Volcánica Esquisto Toba Volcánica o brecha de Toba de color claro Brecha de Toba Volcánica o Pórfido Andesítico Granito, Gneis y Esquisto Ceniza Volcánica o Toba Caliza y Ceniza Volcánica Ceniza Volcánica Pomácea Granito Suave y Gneis Ceniza Volcánica o roca Esquisto Arcilloso Toba Volcánica roca Máfica

Factor	Clasificación
	AGUA Serpentina Lava Máfica o Brecha de Toba Flujo Lodoso o Lahar Máfico roca Serpentina Lava Máfica Material volcánico y Máfico Mezclado Esquisto Arcilloso Calcáreo Arenisca y Esquisto Arcilloso Ceniza volcánica Aluvión Caliza y Aluvión de Esquisto Arcilloso roca Máfica Serpentinizada o Serpentina Serpentina o roca Máfica Serpentinizada Depósitos Aluviales Antiguos o Marinos Sedimentos Depósitos Marinos roca Calcárea Esquisto Arcilloso Calcáreo o Caliza suave Esquisto Calcáreos roca Andesítica roca de Granito y Gneis Intemperizado Toba Volcánica o Brecha de Toba Esquisto o Esquisto Arcilloso Material Máfico Esquisto Arcilloso o Caliza Depósitos Marinos Aluviales Materiales de Grano Fino Ceniza Volcánica o Material Aluvial Corriente de Lodo Material Volcánico Flujo Lodoso Volcánico Materiales Volcánicos Depósitos Marinos o Aluviales Lodo volcánico Arena de Mar Material Sedimentario Arena Volcánica roca Granítica o Ceniza Volcánica Lodo Volcánico Flujo Lodoso Flujo Lodoso Volcánico y Material Aluvial Lahar Máfico Granito, Gneis Materiales Aluviales roca Felsítica
Textura superior	Franca o Franca-Arcillosa

Factor	Clasificación
	Franco-Limosa a Franca Arcillosa Franco-Limosa Franco-Arcillo-Turbosa Franco-Arcillosa Franco-Arenosa Franco-Limosa a Franco-Arcillosa Franco-Arcillo-Arenosa Fina a Franco-Limosa Franco-Arenosa Fina Franco-Arenosa-Gravosa a Franco-Arcillo-Arenosa Franca a Franco-Limosa Arcilla Franco-Arcillosa a Arcilla Franca Franco-Turbosa Franco-Arcillo-arenosa Franco-Arcillo-Limosa Franca a Franco-Arcillosa Franco-Arcillosa o Franca Agua Franca a Franco-Arcillo-Arenosa Arena Fina-Franca Suelta o Arena Muy Fina -Franca Franco-Limosa a Franco-Arenosa Fina o Muy Fina Franco-Arcillo-Arenosa Franco-Gravosa Franco-Arenosa o Gravosa Franco-Arenosa Muy Fina Franco-Arcillo-Esquistosa o Gravosa Franco-Arenosa a Franca Franco-Arcillosa o Franco-Arcillo-Limosa Arcilla o Franco-Arcillosa Franco-Arcillosa a Franco-Arcillo-Micácea Franco-Limosa o Arcilla-Limosa Arcilla-Limosa a Arcilla Arcilla-Limosa Arcilla o Arcilla-Limosa Franco-Arenosa Fina a Franco-Arcillosa o Franco-Arcillo-Limosa Franco-Limosa o Arcillosa Franco-Arenosa o Franco-Arenosa Fina Arena Arcilla-Limosa a Arcilla Pastosa Franca a Franca-Arcillosa Arena-Franca Gruesa o Franco-Arenosa a Arena-Gravosa
Textura superior calidad	Media Moderadamente Fina Muy Fina Agua Gruesa

Factor	Clasificación
	Fina
Textura inferior	Franco-Calcareo o Franco-Arcillosa Franco-Limosa a Franco-Arcillo-Limosa Franco-Limosa o Franco-Arcillo-Limosa a Arcilla Arcilla Franco-Arcillosa Arcillosa Franco-Limosa a Arcilla-Micácea Franco, Arcilloso a Arcilla Franco-Arcillo-Gravosa Arcillo-Arenosa-Gravosa o Arcilla a Franco-Arcillo-Arenosa Franco-Arcillosa a Franco-Arcillo-Micácea Franca o Franco-Arenosa Fina Franco-Arcillosa Gravosa Franco-Arcillo-Limosa Franco-Limosa a Franco-Arcillosa Franco-Arcillosa o Franco-Arcillo-Arenosa Franco-Arcillosa o Arcilla Franco-Arenosa-Fina Franco-arcillosa a Franco-arcillo-limosa Franco-Arcillo-Limosa a Arcilla Franco-Arcillosa a Arcillosa Franco-Arcillo-Arenosa Franco-Arcillosa o Franca Franco-Arenosa Franco-Arenosa a Franco-Arenosa Fina Arcilla a Franco- Arcillosa o Franco-Arcillo- Arenosa AGUA Franco-Arcillosa a Arcilla Ceniza no Intemperizada Arcilla o Franco-Arcillosa Arcilla a Franco-Arcillosa Ceniza Volcánica Franco-Arcillo-Arenosa o Franco-Arenosa Franco-Arenosa Gravosa Franco-Arenosa Fina a Arena-Franca Fina Franco-Arenosa a Franco-Arcillosa Franco-Arenosa, Franco-Limosa o Franco-Arcillosa Franco-Arcillo-Esquistosa o Gravosa Franco-Arcillosa y Esquistosa Franco-Arcillo-Arenosa o Franco-Arcillosa Franco-Arcillosa o Arcilla a Franco-Arcillo-Arenosa (cambiando con la profundidad) Franco-Limosa Franco-Arcillo-Micácea Franco-Limosa a Arcilla Arcilla a Arcilla Gravosa

Factor	Clasificación
	Franco-Arcillo-Limosa a más Liviana Arcilla a Arcillo-Limosa Franca a Franca-Limosa Fina Franca o Franco-Arcillosa Franco-Arcillosa a Franco-Arcillo-Limosa Franca o Franco-Arenosa a Arena-Franca Franco-Arcillosa a Arcilla-Arenosa Arena-Franca Arena Franco-Arenosa o Franca Franca Franco-Arenosa o Franco-Gruesa Franca a Franco-Arenosa Franca a Franca-Arcillosa Arena Gruesa-Franca a Franco-Arenosa Franco-Limosa a Franco-Arcillo-Limoso Arcilla a Arcilla-Arenosa
Textura inferior calidad	Moderadamente Fina Media Muy Fina Agua – Moderadamente Gruesa Fina Gruesa Muy Gruesa
Riesgo Erosión	Alto Regular a Bajo Regular Regular a Alto Muy Alto Bajo Bajo (Alto en las zonas quebradas) Agua Alto (Químicamente) Regular a bajo Bajo (en algunos lugares Regular) Bajo a Regular

Fuente: [22](#)

4.4.1.4. Factores capacidad del uso del suelo

El uso del suelo puede ser un factor importante en el análisis de deslizamientos ya que está relacionado a la actividad humana tales como deforestación, construcciones y cultivos en pendientes altas pueden tener un impacto importante en el deslizamiento.

Los factores de uso del suelo se encuentran divididos en los indicadores en los mapas elaborados y recopilados por el Ministerio de Agricultura de Guatemala a una escala de 1:250000 [22], que pueden observarse en el Cuadro [12]

Cuadro 12: Resumen de propiedades uso de suelo obtenidas del mapa de uso de suelos Guatemala

Factor	Clasificación	Descripción
Capacidad de uso de la tierra	I	Tierras cultivables con ninguna o pocas limitaciones, aptas para la agricultura bajo riego con relieve plano y productividad alto con buen nivel de manejo.
	II	Tierras cultivables con pocas limitaciones, aptas para cultivos bajo riego, relieve plano, ondulado o suavemente inclinado, alta productividad de manejo moderadamente intensivo.
	III	Tierras cultivables, tienen medianas limitaciones para producción agrícola, aptas para cultivos en riego y muy rentables, relieve plano a ondulado o suavemente inclinado, productividad mediana con prácticas intensivas de manejo.
	IV	Tierras cultivables con severas limitaciones permanentes, con relieve ondulado o inclinado, aptas para pastos y cultivos perennes, requieren de prácticas intensivas de manejo.
	V	Tierras preferentemente no cultivables, principalmente aptas para pastos, bosques o para desarrollo de la vida silvestre, factores limitantes muy severos para cultivos, generalmente drenaje y pedregosidad, con relieve plano u ondulado.
	VI	Tierras no cultivables, principalmente para producción forestal. Factores limitantes muy severos de relieves, profundidad y rocosidad.
	VII	Tierras no cultivables, aptas solamente para fines de producción forestal, relieve quebrado con pendientes muy inclinadas.
	VIII	Tierras no aptas para el cultivo, solo aptas para parques nacionales, recreación y vida silvestre y para protección. Relieve muy quebrado, escarpada o playones.
	Agua	

Fuente: [22]

4.4.1.5. Factores fisiográficos

Cuadro 13: Resumen de propiedades fisiográficas obtenidas del mapa de fisiografía de Guatemala

Factor	Clasificación
Fisiografía nombre	Tierras Altas Sedimentarias Tierras Altas Cristalinas Tierras Altas Volcánicas AGUA Depresión de Izabal Depresión del Motagua Tierras Bajas Interiores de Petén Plataforma Sedimentaria de Yucatán Cinturón Plegado del Lacandón Montañas Mayas Pendiente Volcánica Reciente Llanura Costera del Pacífico

Fuente: [22]

4.4.2. Factores disparadores

Los factores desencadenantes son los factores dinámicos que pueden causar la iniciación del movimiento y aumentar la fuerza impulsora. Estos factores incluyen lluvias, terremotos y erupciones volcánicas.

4.4.2.1. Factor precipitación

El factor de precipitación podría ser el más importante de los factores por ser el más constante y sabiendo que el agua es el peor enemigo de un deslizamiento dado su efecto de acumulación en el suelo modificando la fuerza y estructura del suelo, puede terminar de romper la estabilidad y producir el deslizamiento [11]. Los catálogos de Google Earth Engine que nos interesan para la generación de precipitación media son los siguientes:

- El mapeo satelital global de precipitación (GSMaP) proporciona precipitación global por hora con una resolución de 0,1 x 0,1 grados. GSMaP es un producto de la misión Global Precipitation Measurement (GPM), que proporciona observaciones de precipitación global. El algoritmo de recuperación de la tasa de precipitación GPM se basa en un modelo de transferencia radiativa. La tasa de nivel ajustada se calcula en función de la optimización de la acumulación de 24 horas de la tasa de lluvia por hora de GSMaP al nivel de precipitación diaria medido por NOAA/CPC [21].

En el Cuadro [7] se puede ver el valor hourlyPrecipRate que es tomado del satélite GMap para la obtención de la precipitación acumulada por hora, para posteriormente hacer una suma de los últimos 7 días.

4.4.3. Inventario de deslizamientos (base de datos)

El factor de inventario de derrumbes se podría ver como el más importante para reconocer insights de un historial de los derrumbes logrando reconocer áreas y posiciones de derrumbes, los factores causantes, las condiciones previas al desastre y también recopilar el daño causado [19]. Por ello, lo mínimo necesario en cada derrumbe guardado en el inventario debería ser la posición en la que ha ocurrido el derrumbe, la fecha del derrumbe y la mayor cantidad de datos posibles sobre los factores ambientales y disparadores de aquel momento. Es importante reconocer también que este inventario de derrumbes debe ser actualizado constantemente y entre mayor sea la cantidad de derrumbes al igual que el tiempo considerado en el historial será más útil para evaluar zonas de riesgo y susceptibilidad de derrumbes. Todo ello conlleva un cuidadoso análisis de modelos de terrenos y correlaciones de campo con los mapas digitalizados [24]. En este caso para la recopilación de todos los factores necesarios, ya los se han evaluado con anterioridad en esta sección teniendo que obtener los factores geomorfológicos y de precipitación por medio de imágenes satelitales obtenidas de Google Earth Imagery [11], mientras que los demás factores por mapas digitalizados de la región de estudio de suelos, geología, fisiografía y uso del suelo utilizando herramientas GIS.

4.5. Vehículos aéreos no tripulados

En la última década, han emergido tecnologías pequeñas, baratas y lo suficientemente ligeras como para permitir la observación aérea. Actualmente, aeronaves con tamaños reducidos son capaces de volar y recopilar información con una mínima intervención humana; estas pequeñas aeronaves se denominan vehículos aéreos no tripulados, o UAV y cuentan con una amplia variedad de tamaños, costos y resistencias [25].

Los UAV son una buena representación de la automatización. A partir de la computación de bajo consumo, cámaras, posicionamiento, transmisión de datos, sensores y baterías, los UAV brindan una manera eficiente de capturar imágenes e información [25]. En la clasificación y elección de un UAV influyen factores superficiales como el tamaño, el peso y la fuente de poder y de factores de diseño como las alas, el multirrotor y el nivel de automatización, sin embargo, el tipo de UAV depende del sistema y de la tecnología que permita que el UAV vuele, los principales sistemas son el de ala fija y el sistema multirrotor aunque también existen aquellos que son tanto multirrotor como de ala fija [26].

4.5.1. Tipos de UAV

Los sistemas de ala fija usan alas fijas y estáticas en combinación con la velocidad aerodinámica de avance para generar el levantamiento del UAV [18]. Aunque es un tipo de UAV, el término ala fija proviene de la industria de aviación, un ejemplo de este sistema son los aviones tradicionales y diferentes tipos de planeadoras como alas delta o parapentes. Por otro lado, los sistemas multirrotor, con un funcionamiento similar al de los helicópteros, dado que posee alas giratorias que las utiliza para generar el levantamiento [19]. Los sistemas rotativos casi siempre están equipados con múltiples rotores pequeños, que son necesarios

para su estabilidad, de ahí el nombre “multirotor” [26].

Figura 18: UAV de ala fija



Fuente: [26]

Figura 19: UAV multirotor



Fuente: [26]

Una característica importante de los UAV de ala fija, es que mayoría de los UAV de ala fija tienen un tiempo de vuelo de varias horas, algunos pueden volar incluso hasta por 16 horas, sin embargo, la mayoría de estos no pueden ser usados para fotografía aérea donde se requiere que el UAV permanezca inmóvil durante un periodo de tiempo. Por otro lado, los sistemas multirotor son usualmente utilizados para toma de fotografías o vídeos, sin embargo, no son apropiados para la cartografía aérea de larga distancia o el reconocimiento debido a que tienen tiempos de vuelo limitados [27].

4.5.2. Nivel de automatización

Muchos de los UAV en la actualidad carecen de piloto, es decir, que ya cuentan con un nivel de automatización para poder realizar viajes programados. El Departamento de Defensa de los Estados Unidos distingue cuatro niveles de autonomía en sistemas no tripulados como son los UAV, el nivel más básico es un sistema operado por humanos en el que un operador toma todas las decisiones al momento del vuelo. En el segundo nivel se encuentran los sistemas delegados, los cuales pueden realizar funciones independientemente del control

humano y realizan tareas cuando se le delega hacerlo sin intervención humana [26].

Un sistema supervisado por humanos, comprendería el tercer nivel, este sistema puede realizar varias tareas cuando un ser humano le otorga ciertos permisos e instrucciones y tanto el sistema como el supervisor pueden tomar acciones basándose en lo detectado. En el cuarto y más alto nivel de automatización, es un sistema completamente autónomo, recibe comandos ingresados por un humano y traduce estos comandos en tareas específicas sin más interacción humana [26].

4.5.3. Fuentes de energía

Una característica muy importante de los UAV es la fuente de energía, y actualmente existen 4 tipos importantes de fuentes de energía que son combustible de avión tradicional, celdas de batería, celdas de combustible y celdas solares [26].

El combustible de avión se utiliza principalmente en UAV de gran tamaño de ala fija, como por ejemplo el UAV militar Predator. Las celdas de batería se utilizan principalmente en UAV multirrotor que por lo general son más pequeños. Estos UAV tienen un menor alcance y requieren un menor tiempo de operación. Un ejemplo de UAV con celdas de batería es el UAV Phantom [26].

Una celda de combustible es un dispositivo electroquímico que convierte la energía química del combustible directamente en energía eléctrica. Actualmente, este tipo de energía es muy poco utilizada en UAV. Solo los UAV de ala fija pueden equiparse con esta celda debido al peso relativamente alto de la celda. Una gran ventaja de usar una celda de combustible es el hecho de que los UAV pueden volar distancias más largas sin recargar. Un ejemplo de un UAV de este tipo es el UAV Stalker y que tiene un tiempo de vuelo de 8 horas en lugar de 2 horas, que es lo más común para un UAV [26].

Por último, están los UAV que utilizan celdas solares y que al igual que los UAV con celdas de combustible, son raros en la industria de UAV actual. Estos son principalmente UAV de ala fija, debido a la baja eficiencia de la corriente. Sin embargo, las celdas solares son adecuadas para pequeños ornitópteros [26].

4.6. Modelos de *machine learning*

4.6.1. *Machine learning*

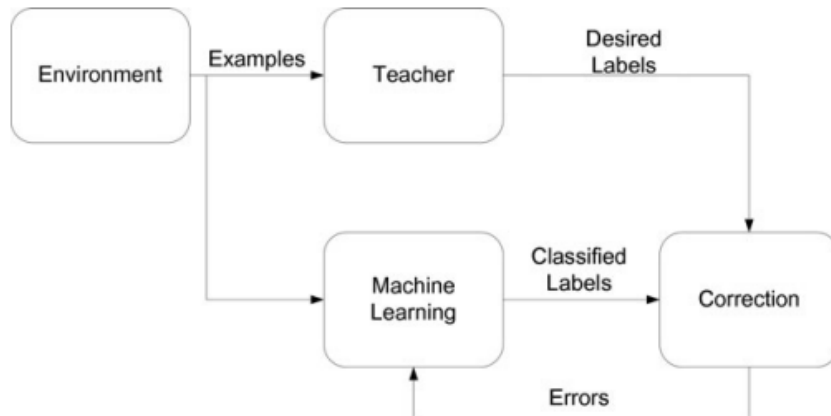
Machine learning es definido como el paradigma de computación donde la capacidad para resolver un problema es construida por ejemplos anteriores [28]. Es decir, basar el razonamiento de una decisión basado en casos similares que han producido precedentemente. Los ejemplos utilizados para construir esta capacidad son los llamados datos de entrenamiento y el proceso es llamado aprendizaje. Luego, de que se ha aprendido, el proceso de resolver el problema es llamado generalización.

Dentro de *Machine Learning*, se encuentran los problemas de clasificación que tratan,

como su nombre lo dice, de clasificar objetos entre una o varias categorías predefinidas y, por otra parte, los problemas de regresión que estiman un valor o varios valores continuos. El aprendizaje supervisado es el primer tipo de *Machine Learning* y se basa en asumir que todos los ejemplos anteriores con los que se realiza un entrenamiento se encuentran etiquetados sobre su resultado, mientras que el entrenamiento no supervisado es otro paradigma en el que los ejemplos de entrenamientos no están etiquetados y su aprendizaje se basa en agrupar por similitudes sin realmente poder dar una etiqueta.

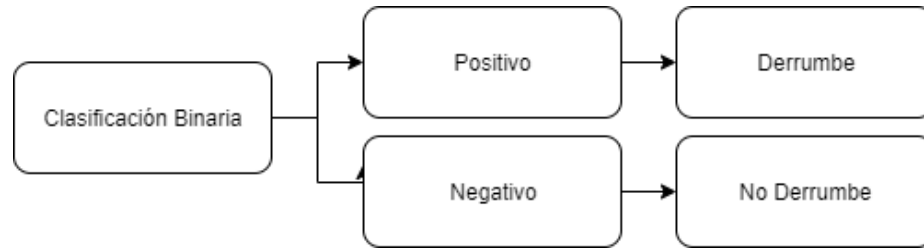
Las aplicaciones de aprendizaje supervisado son enormes y pueden encontrarse en miles de ejemplos de la vida diaria en el que se quiere encontrar una relación entre varias características para obtener una predicción. En la Figura 21 se puede observar el funcionamiento básico de aprendizaje supervisado. Se debe considerar que existen dos valores de salida en este aprendizaje, el "target output" asignado a cada ejemplo de entrenamiento y el "computed target" que ha predicho el algoritmo de aprendizaje. Lo que se intenta es minimizar la diferencia entre el "target output" y el "computed output" de forma los parámetros utilizados sean optimizados y finalmente, minimizar el error de una clasificación incorrecta. Cada ejemplo, cuenta con un set de "features o parameters", que serán utilizadas para realizar la clasificación. El proceso es iterativo para minimizar el error y encontrar una convergencia de los "parameters". Para el caso de los derrumbes, se pueden centrar en un aprendizaje supervisado de clasificación ya que nuestro label es si existió un derrumbe o no, siendo una clasificación binaria (dos categorías posibles, ver Figura 20) a partir del inventario de derrumbes de Guatemala (ejemplos de entrenamiento).

Figura 20: Diagrama de clasificación binaria de derrumbes



Fuente: elaboración propia

Figura 21: Diagrama de aprendizaje supervisado



Fuente: [28]

4.6.2. Preprocesamiento del *training set*

Para un buen funcionamiento de nuestros algoritmos de clasificación, se debe asegurar varios aspectos de nuestra *data* de ejemplos, también llamada "*training set*". Es importante conocer que los algoritmos de *Machine Learning* que la *data* sea estructura en vectores numéricos dado que los principios de los mismos se encuentran en las matemáticas. Por ello, se puede listar las siguientes tareas necesarias para el preprocesamiento del *training set* a tomar en cuenta [29]:

- **Data Encoding:** Es necesario convertir las *features* categóricas en un valor numérico. Para ello, se utiliza *Text Encoding*, lo cual permite convertir variables de tipo texto a tipo numérico. Los valores categóricos son mapeados hacia valores de enteros, para luego poder ser representado por este valor.
- **Data Scaling:** Algunos algoritmos, como lo son las regresiones, necesitan que la *data* numérica se encuentre escalada y normalizada ya que si no se hiciese este, surgirían errores de dar mayor peso a ciertas *features* apriori del entrenamiento. Para ello, se utiliza un Standard Scaler cuando el algoritmo lo amerite.
- **Inconsistent Data:** Al analizar el *training set* podría llegar a suceder que en ciertas *features* existan valores inconsistentes a lo que se plantea. Por ello, siempre es óptimo revisar los datos de nuestro *data set* y hacer una exploración que permite eliminar estos datos si son pocos o ya bien buscar una estrategia para reemplazarlos si fuese una cantidad alta.
- **Missing Feature Values:** Para algunas *features* del dataset, podría ser que existan datos faltantes para uno o varios ejemplos de entrenamiento. Dependiendo del algoritmo se debe de tomar en cuenta, estrategias para resolver estos datos faltantes.
- **Irrelevant features:** En cada dataset, debe considerarse que las *features* tomadas en cuenta sean relevantes para el caso de estudio que se realiza. Esto puede hacerse por medio de investigación previa respecto al tema que se desea predecir y la validación de expertos. La toma de *features* irrelevantes causa ruido alrededor de la predicción empeorando los resultados.
- **Feature Values Noise:** Al igual que los *features* irrelevantes, pueden existir *features* que tengan datos que causen ruido al algoritmo. Esto quiere decir, que los valores con-

tenían errores, fueron medidos pobremente, fueron obtenidos de fuentes no confiables, etc, por lo que se debe de igual forma validar estos datos con expertos.

4.6.3. Algoritmos de clasificación

A continuación se presenta una descripción de los algoritmos de clasificación: *Decision trees*, *Random forest*, *XGBoost* y *Support vector machine*.

4.6.3.1. *Decision tree* y *Random forest*

4.6.3.1.1. *Decision tree*

Los *Decision tree* (árboles de decisión) utilizan árboles creando ramas de decisión hasta llegar a nodos terminales que realizan la clasificación en las correspondientes categorías predefinidas. Durante la etapa de aprendizaje, el árbol se construye a sí mismo sobre los ejemplos de entrenamiento. Este algoritmo es uno de los modelos no lineales más simples de *Machine Learning* [30].

El proceso de construcción es intuitivo ya que puede verse como varias condicionales anidadas que permiten finalmente llegar una predicción. Por lo general, el algoritmo cuenta de 2 pasos base: *Splitting* y *Prunning*.

1. ***Splitting***: Esta parte incluye el proceso de tomar la decisión de un nuevo "split" en el árbol. Puede existir un número gigante de posibles "split", por lo que se utiliza una métrica como el lowest mean squared error, factor de entropía o factor de gini para tomar la decisión de cuál es el mejor split que se podría tomar en el árbol en dicho momento. Luego de este split, se obtiene la separación de los datos iniciales del nodo en otros 2 grupos. Por lo tanto, se debe repetir este proceso hasta que se obtenga un punto singular que no pueda dividirse o ya bien si se ha llegado a la estrategia de paro que es el siguiente paso (prunning) [30].
2. ***Prunning***: Esta segunda parte intenta que el modelo reduzca su complejidad y que en la medida de lo posible no exista un *overfit* de datos. Queriendo decir que a pesar de que es posible, dejar que el árbol, se divida hasta no poder más, no siempre es lo mejor. La estregia de "prunning" es cortar las ramas del árbol menos necesarias para evitar *overfitting* o ya bien a priori darle un máximo de profundidad al árbol del que puede explorar.

Cuadro 14: Parámetros importantes de *decisiontreeclassifier* en librería scikit-learn de python

Parámetro	Descripción
Criterio	La función para medir la calidad del <i>split</i> , pudiendo ser gini, entropy o logloss.
<i>Max depth</i>	La máxima profundidad del árbol.
<i>Max features</i>	La mayor cantidad de <i>features</i> a considerar para encontrar el mejor <i>split</i> .

Fuente: elaboración propia

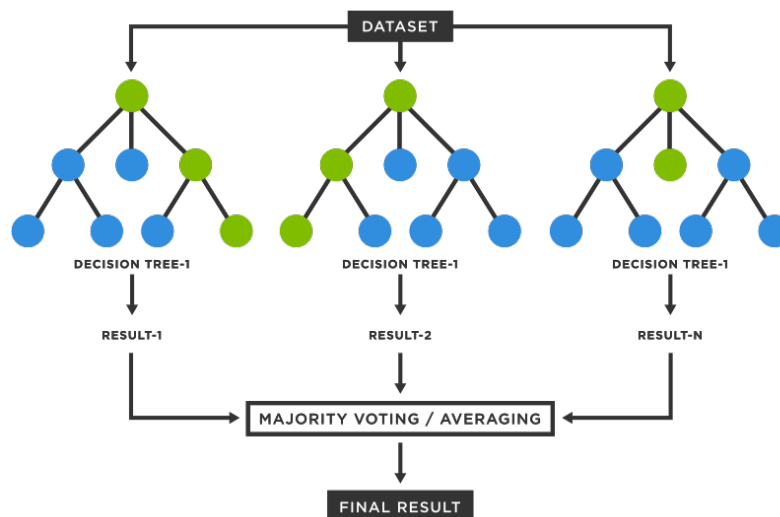
Los puntos más importantes a considerar sobre este algoritmo:

- Los grandes beneficios de este algoritmo son la interpretabilidad [29] ya que claramente se puede visualizar cómo funciona el algoritmo y en qué decisiones se basa para finalmente devolver un resultado de clasificación
- Se debe tomar en cuenta que con los parámetros incorrectos se tenderá a realizar un *overfit* de la *data*, sucediendo que es muy bueno para los datos de entrenamiento, pero muy malo para generalizar que es lo que desea con *Machine Learning*.
- De un mismo *training set*, pueden existir varios árboles de decisión y como regla deben preferirse los árboles más pequeños ya que tienen menor probabilidad de *overfitting* y mayor interpretabilidad [28].

4.6.3.1.2. *Random forest*

Random forest se basa en la idea de particionar el *training set* en varios subconjuntos que permitan construir diferentes árboles de decisión por cada *subset*. Finalmente, la clasificación se da por la votación del *output* de cada árbol como puede observarse en la Figura 22. El nombre *Random forest* se ilustra con que se basa en varios árboles de decisión, sumado a que cada *subset* es una partición aleatoria del *training set* [28]. Al incluir varios árboles para la toma de decisión se agrega complejidad al modelo y de igual manera se mejora el *performance* basado en la idea de "conocimiento colectivo" llamado *Ensemble Learning*.

Figura 22: ¿Cómo funciona *Random forest*?



Fuente: [31]

Ensemble Learning tiene dos métodos para crear los "ensembles", siendo el primero "bagging" y el segundo "boosting". *Random forest* se basa en el concepto de *bagging* (*Bootstrap Agreggation*), que tiene dos partes importantes [30]:

1. **Bootstrap:** Esto es la idea que cada árbol de decisión aprenda un subconjunto ligeramente distinto de los otros. Para cada árbol de aprendizaje individual, se genera una *data* por un proceso de remuestreo del *training set* original (notando que en *Random forest* se realiza con reemplazo).
2. **Aggregation:** La agregación es el proceso que permite utilizar múltiples árboles al ejecutarse el *bootstrap* múltiples veces teniendo una cantidad *n* de estimadores. Finalmente, se genera una gran cantidad de árboles de decisión ligeramente distintos logrando que el error total se logre promediar y disminuir al contar con varios aprendices que combinan sus decisiones como votos para una decisión final.

Cuadro 15: Parámetros importantes de *randomforestclassifier* en librería *scikit-learn* de python

Parámetro	Descripción
n estimators	La cantidad de árboles en el <i>forest</i> .
Criterion	La función para medir la calidad del <i>split</i> , pudiendo ser gini, entropy o log_loss.
Max depth	La máxima profundidad del árbol.
Max features	La mayor cantidad de <i>features</i> a considerar para encontrar el mejor <i>split</i> .

Fuente: elaboración propia

Los puntos más importantes a considerar sobre este algoritmo:

- *Random forest* es un tipo de aprendizaje de *Ensemble Learning* basado en la estrategia de *Bagging*.
- El parámetros más importantes de este algoritmo son la cantidad de árboles en el *forest* que serán los votantes para el resultado final.
- La importancia de las variables es la forma de interpretar los resultados del *Random forest* y puede ser clave para seguir mejorando el modelo.

4.6.3.2. *AdaBoost, XGBoost y LightGBM*

AdaBoost, XGBoost y *LightGBM* son otro tipo de algoritmos basados en árboles de decisión y *Ensemble Learning*. No obstante, cambian el método que utiliza su *Ensemble Learning*. Para *Random forest*, se ha visto que utilizaba *bagging* y aunque puede producir resultados muy buenos, sufre del problema que cada clasificador individual ha sido independiente uno del otro. Por este motivo, existe el approach de "*boosting*", el cual se basa en que cada clasificador individual complementa a los demás [29].

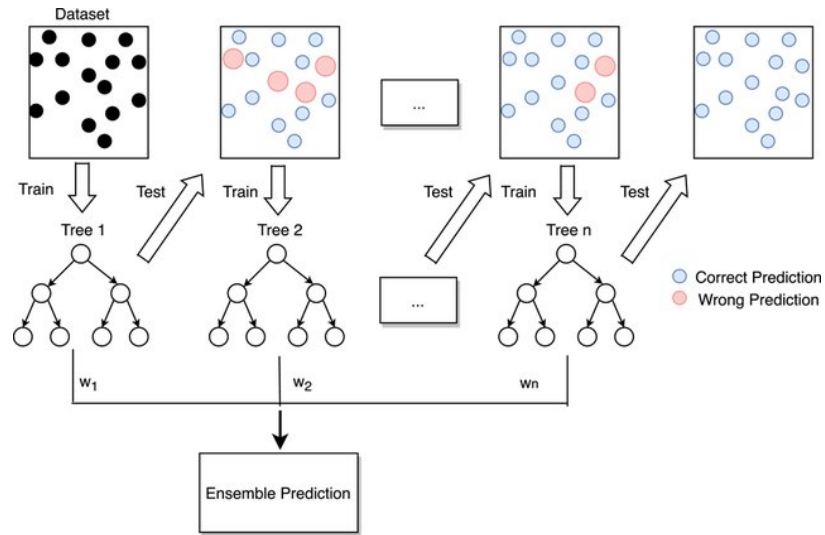
Boosting, al igual que *Bagging*, crea árboles de decisión ligeramente diferente entre ellos por medio de un proceso iterativo, haciendo más énfasis en los aprendizajes débiles. En cada iteración, cada punto individual del *set* se le es otorgado un peso, logrando que cada nuevo clasificador haga mayor énfasis en aprender las cosas que aún no han sido entendidas por los clasificadores pasados y conclusivamente, mejor el ensemble [30]. En la Figura [23], puede observarse el flujo que es seguido por el método de *Boosting*. Este proceso iterativo es llamado *Gradient Boosting*, basado en la matemática de que el término gradiente se refiere al vector de derivadas parciales que apuntan en la dirección de la pendiente más pronunciada. Cada vez que se agrega un árbol, se necesita agregar un árbol que mejor explique las variaciones que aún no han sido explicadas por los árboles anteriores, encontrando un resultado por medio de la optimización en derivadas parciales.

Cabe resaltar, que al comienzo no eran muy buenos en *performance*, pero que hoy en día existen varios approach sobre *Gradient Boosting Algorithms*, siendo el primero de ellos *AdaBoost* y *XGBoost* y *LightGBM* unos de los más conocidos y utilizados popularmente ahora por sus buenos resultados.

Los puntos más importantes a considerar sobre este algoritmo:

- *Gradient Boosting* es un método de *Ensemble Learning* que se basa en aprender iterativamente lo que aún no ha sido entendido por el modelo, logrando mejores resultados que *bagging* [30].

Figura 23: Diagrama del flujo del *boosting method*



Fuente: [32]

4.6.3.2.1. *AdaBoost*

AdaBoost fue el primer algoritmo de GradientBoosting eficiente a adaptarse a "weak learners". La idea general es que utiliza distribución probabilística creando cada clasificador uno por uno. Cada ejemplo tiene una probabilidad de ser escogido y la suma es 1; sin embargo, cada ejemplo tiene un peso diferente, dando mayor peso a los ejemplos mal clasificados, logrando que tengan una mayor probabilidad de ser escogidos en la siguiente iteración. Finalmente, en comparación de *bagging* cada clasificador también tiene un peso respecto a la votación, por lo que sea realiza una votación por mayoría ponderada [29].

Cuadro 16: Parámetros importantes de *adaboostclassifier* en librería scikit-learn de python

Parámetro	Descripción
n estimators	La cantidad de árboles en el que el <i>boosting</i> es terminado.
<i>Learning rate</i>	Peso aplicado a cada clasificador luego de cada iteración de <i>boosting</i> . Una mayor <i>Learning rate</i> aumenta la contribución de cada clasificador.

Fuente: elaboración propia

4.6.3.2.2. *XGBoost*

XGBoost es uno de los algoritmos más utilizados en *Machine Learning* por su alto rendimiento y *performance*. *XGBoost* utiliza un "histogram-based splitting", que quiere decir que en vez de iterar sobre todos los posibles *splits*, el modelo construye histogramas de cada una de las variables y las utiliza para encontrar el mejor *split* por variable [30]. Además, utiliza un *fit tree-wise*, dando un poco de menor *accuracy* pero dejando menor espacio a *overfitting*.

Cuadro 17: Parámetros importantes de *xgbclassifier* en librería *xgboost* de python

Parámetro	Descripción
n estimators	La cantidad de árboles en el que el <i>boosting</i> es terminado.
<i>Learning rate</i>	Peso aplicado a cada clasificador luego de cada iteración de <i>boosting</i> . Una mayor <i>learning rate</i> aumenta la contribución de cada clasificador.
<i>Max leaves</i>	El número máximo de hojas
<i>Max depth</i>	La máxima profundidad del árbol.
<i>Max features</i>	La mayor cantidad de <i>features</i> a considerar para encontrar el mejor <i>split</i> .

Fuente: elaboración propia

4.6.3.2.3. LightGBM

LightGBM es un algoritmo inventado por Microsoft menos popular que *XGBoost*, pero que ha ganado relevancia en los últimos años, mejorando en velocidad y uso de memoria. La diferencia con *XGBoost* es que tiene un método más efectivo para definir *splits* llamado "Gradiente-Based One-Side Sample (GOSS)". GOOS computa gradiente para cada punto de la *data* y deja afuera los puntos con un gradiente bajo ya que estos ya han sido comprendidos bien por clasificadores anteriores [30]. Por último, también utiliza un *fit leaf-wise* y un approach de "Exclusive Feature Bundling", que permite dar mayor velocidad al aprendizaje cuando se tienen muchas variables correlaciones de las cuales escoger.

Cuadro 18: Parámetros importantes de *lgbclassifier* en librería *lightgbm* de python

Parámetro	Descripción
n estimators	La cantidad de árboles en el que el <i>boosting</i> es terminado.
<i>Learning rate</i>	Peso aplicado a cada clasificador luego de cada iteración de <i>boosting</i> . Una mayor <i>learning rate</i> aumenta la contribución de cada clasificador.
Num <i>leaves</i>	El número máximo de hojas
<i>Max depth</i>	La máxima profundidad del árbol.
<i>Max features</i>	La mayor cantidad de <i>features</i> a considerar para encontrar el mejor <i>split</i> .

Fuente: elaboración propia


4.6.3.3. Métodos probabilísticos y estadísticos: *Logistic regression* y *Support vector machine*

4.6.3.3.1. *Logistic regression*

Logistic regression es un modelo matemático utilizado la predicción de una clasificación

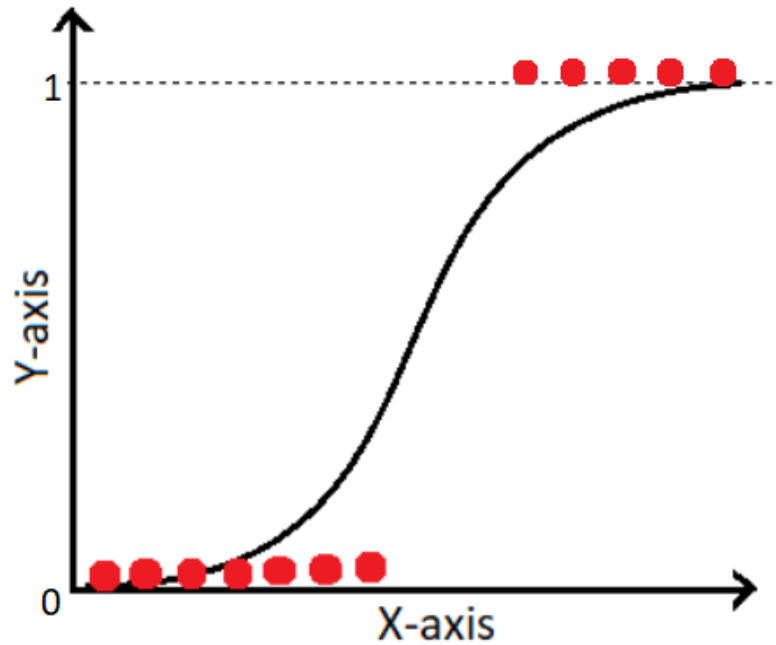
categoría. El problema que intenta resolver es el problema multivariable que busca encontrar la relación entre una o más variables de estudio y una salida binaria que representa una etiqueta positiva o negativa [33]. Las primeras son llamadas las variables independientes, mientras que la salida es la variable dependiente. *Logistic regression* es una modelación matemática oportuna para resolver estos problemas. En general, los parámetros de este algoritmo no suelen cambiarse, y suele evaluarse los coeficientes generados del modelo por cada variable. Si el coeficiente es positivo, cuando la variable aumenta, la probabilidad de ser una etiqueta positiva también lo hace, mientras que si el coeficiente es negativo, la probabilidad de ser una etiqueta positiva disminuye.

Figura 24: *Logistic model*

$$z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

$$f(z) = \frac{1}{1 + e^{-z}}$$
$$= \frac{1}{1 + e^{-(\alpha + \sum \beta_i X_i)}}$$

Fuente: [33]

Figura 25: Gráfica de *logistic regression*



Fuente: [33](#)

4.6.3.3.2. *Support vector machine*

Support vector machine es definido como un clasificador dual de hiperplano [28](#). Cuando se tiene un alto dominio es un algoritmo muy útil, con una complejidad matemática muy grande y, por lo tanto, puede ser computacionalmente costoso. Este se encuentra basado en la ecuación matemática de un clasificador lineal $y = wx' + \gamma$. El espacio dual se asume al crear un "feature space" a partir del espacio original, convirtiendo un espacio de no separabilidad lineal hacia un espacio con separabilidad lineal. El proceso de aprendizaje pretende construir los hiperplanos duales con el máximo margen de separación en el *feature space*. El mapeo explícito de ejemplos de entrenamiento se evita al utilizar el *kernel function* visto como el producto interno de 2 vectores en el *feature space*.

Cuadro 19: Parámetros importantes de *supportvectorclassification* en librería *sklearn* de python

Parámetro	Descripción
C	Parámetro de regularización.
<i>kernel</i>	Especifica el <i>kernel</i> utilizado por el algoritmo, pudiendo ser lineal, poly, rbf, sigmoid y precomputed.
<i>degree</i>	Define el grado polinomial si el <i>kernel</i> es poly.

Fuente: elaboración propia

4.6.3.4. Evaluación y validación del *performance* los algoritmos

4.6.3.5. Evaluación del modelo y validación

Para realizar la evaluación de un modelo, por lo general se divide en 3 categorías, el dataset. [34].

1. **Training Data:** Esta es la *data* que se utiliza para entrenar el modelo.
2. **Testing Data:** Esta es la *data* que el modelo entrenado no ha visto antes, y, por lo tanto, se puede evaluar el *performance* del *test*.
3. **Validation Data:** Esta *data* no siempre es separada por cuestiones de limitaciones de *data* o tiempo. El propósito debería ser chequear si el modelo está realizando *overfitting*.

Otra forma de evaluación es el *K-Fold Cross-Validation*. Esta técnica, en comparación de dividir nuestra *data* en *train* y *test*, se basa en muestreo probabilístico para validar el modelo.

El *performance* de un modelo de clasificación puede evaluarse por las siguientes métricas [34]:

- **Accuracy:** Es la proporción de las predicciones correctas.
- **Confussion Matrix:** En la matriz de confusión se puede evaluar las accuracies del modelo por cada una de las clases a predecir. Por ello, se puede obtener los *True Positive*, *False Positive*, *False Negative* and *True Negative*.
- **Sensitivity:** Es la probabilidad de que al predecir una clase como positiva sean realmente positiva. También llamada *True Positive rate TPR* o *precision*.
- **Specificity:** Es la probabilidad de que al predecir una clase como falsa sean realmente falsa. También llamada *True Negative rate TNR* o *recall*.
- **Area under ROC Curve:** Es la representación gráfica de un clasificador binario mostrando la relación entre *TPR* y *TNR*, en dónde se intenta maximizar ambas medidas.

Además, de estas medidas se deben tomar en cuenta las medidas del Cuadro [5] en el que intersectan algunas para la validación del modelo de susceptibilidad de derrumbes.

4.6.3.6. *Hyperparameter optimization*

Hyperparameter optimization es el proceso de la selección de los mejores *Hyperparameters* para un algoritmo de *Machine Learning*. El *set* de *Hyperparameters* es necesario porque maximizan el *performance* del modelo. Los métodos más utilizados son los siguientes [34] y se pueden definir como procesos iterativos de búsqueda de los parámetros óptimos:

- **Manual Search:** Buscar parámetros de forma manual y con juicio de experiencia.
- **Manual y Automatic Grid Search:** Buscar parámetros de forma manual o automática de acuerdo a un grid.
- **Random Search:** Buscar parámetros de forma aleatoria que encuentren los mejores parámetros. Esto es útil porque ahorra espacio y tiempo, pero no siempre se encuentran los óptimos.

La optimización de los *Hyperparameters* puede conducir a un *overfit*, lo que a menudo resulta en que un modelo se vuelva demasiado complejo. Por lo tanto, es importante utilizar datos en conjuntos muestreados y métricas apropiadas para la evaluación.

4.7. Segmentación semántica de imágenes

La segmentación semántica de imágenes es un proceso esencial en la visión por computadora para dividir una imagen original en regiones significativas no superpuestas, por lo que cada pixel de la imagen corresponde a solo un tipo específico. Durante las últimas décadas, la segmentación de imágenes ha sido una tarea muy desafiante para la visión por computadora. Esta difiere de clasificación de imágenes o reconocimiento de objetos dado que no es necesario saber que son los objetos visuales de antemano [35].

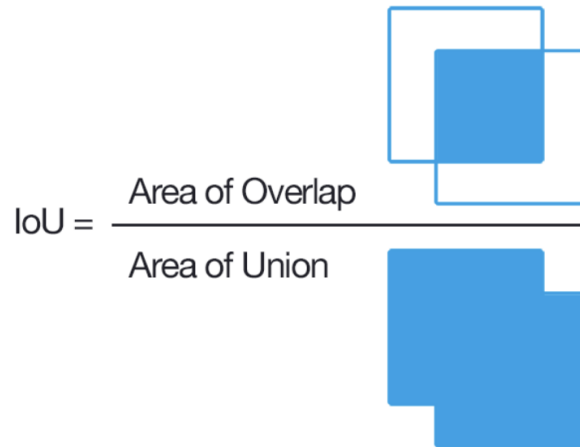
La segmentación de imágenes tiene una variedad de aplicaciones, como la detección de copia por imagen, análisis de imágenes tomadas por UAV, segmentación de objetos en imágenes satelitales. En todos estos ejemplos, tener imágenes segmentadas y aplicando un algoritmo de segmentación de imágenes ayudaría a abordar el problema a un nivel semántico [35].

4.7.1. Métricas para modelos de segmentación semántica de imágenes

Dado la complejidad de los modelos de visión por computadora, existen métricas específicas para evaluar la efectividad de estos modelos. En el caso de la detección y segmentación de objetos, existe la métrica Intersection Over Union (IoU). Para el caso de la detección de objetos, IoU es una métrica auxiliar, pero en el caso de la segmentación semántica de imágenes, la métrica IoU se utiliza como la métrica principal de evaluación de los modelos.

IoU es un número que cuantifica el grado de superposición entre dos cajas, que serían la máscara real de una imagen y la predicción de dicha máscara. Como se puede ver en la Figura [26], el valor obtenido de esta métrica es la división entre el área de superposición entre las máscaras, con el área de unión de las dos máscaras.

Figura 26: IoU



Fuente: [35](#)

4.7.2. Algoritmos para segmentación semántica de imágenes

Los algoritmos de *deep learning* han ayudado a solucionar muchos de los retos que presentan las tareas más complejas de visión por computadora. En la actualidad existen diferentes modelos que han sido diseñados para justamente estos problemas de visión por computador, como lo es la segmentación semántica de imágenes.

4.7.2.1. U-Net

La U-Net es una modificación de una red neuronal totalmente convolucional y fue creada por motivos médicos en 2015 por Olaf Ronneberger [36](#). La arquitectura de esta red permite que esta capture información más precisa y retenga más información mediante la concatenación de características de alto nivel con las de bajo nivel. Pero este proceso de concatenar la información de varios bloques bloquea a la U-net a producir resultados más precisos.

4.7.2.2. Feature Pyramid Network (FPN)

El modelo FPN es una red neuronal totalmente convolucional y fue diseñada por Tsung-Yi Lin [37](#) en 2016 y es utilizada para detección de objetos y para segmentación semántica de imágenes. La arquitectura de esta se compone de un camino de abajo hacia arriba, otro de arriba hacia abajo y conexiones laterales que unen características de baja y alta resolución. Este modelo es eficiente porque de esta forma se propaga mejor la información baja en la red.

4.7.2.3. *Pyramid Scene Parsing Network (PSPNet)*

LA *PSPNet* o red de análisis de escenas piramidales, fue creada en 2016 por Hengshuang Zhao [38] y es una red eficaz para la comprensión de escenas complejas y tener un entendimiento completo de la imagen o escena analizada. Esto se logra a través de que la red aprenda mejor la representación del contexto global de una escena. Los patrones de la escena se extraen de la imagen de entrada mediante un extractor de características con una estrategia de red dilatada. Esta red neuronal se utiliza mayormente para segmentación de imágenes con muchos objetos de formas y tamaños variados.

4.7.2.4. *Path Aggregation Network (PANet)*

Path Aggregation Network (PANet) es un modelo creado por Shu Liu en el año 2018 [39]. Esta red se enfoca en mejorar la propagación de la información de una imagen. El extractor de características de la red utiliza una arquitectura FPN con una nueva ruta de abajo hacia arriba aumentada que mejora la propagación de características de capa baja.

4.7.2.5. *DeepLab V3+*

Google desarrolló Deeplab que usa como primera estructura una red neuronal convolucional (CNN). Actualmente, existen 3 versiones de Deeplab que han ido mejorando en cada versión. La última versión *DeepLab V3* incluye la arquitectura encoder-decoder. Las ventajas de implementar esto son que puede capturar límites de objetos de diferentes formas y además captura información de alta semántica, dado que los codificadores reducen gradualmente la imagen de entrada mientras extraen información espacial vital [40].

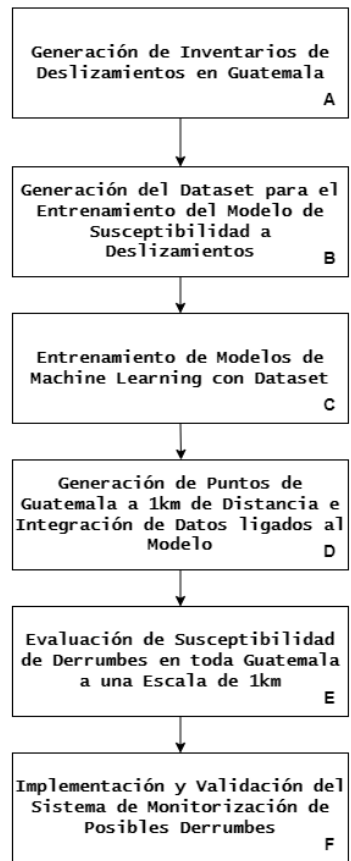
5.1. Metodología general

Este megaproyecto tuvo como principal objetivo el diseñar e implementar un sistema de *software* de monitoreo de derrumbes para zonas de riesgo, de modo que se pudiese visualizar en el mapa de Guatemala, un grado de susceptibilidad de deslizamientos. Para lograr esto se realizó una investigación sobre bibliografía ya existente que enfrente este problema, así como sobre los conceptos más importantes de la ingeniería civil que influyen en la incidencia de estos fenómenos. Para finalmente diseñar e implementar el sistema de monitoreo de susceptibilidad de desplazamientos.

Luego de esta investigación previa se dividió el proyecto en seis grandes procesos, como se puede ver en la Figura [27](#), los cuales fueron: la generación de inventarios de deslizamientos en Guatemala, la generación del dataset para el entrenamiento de modelo de susceptibilidad a deslizamientos, el entrenamiento de modelos de *machine learning* con el dataset, la generación de puntos de Guatemala a 1km de distancia e integración de datos ligados al modelo, la evaluación de susceptibilidad de derrumbes en toda Guatemala a una escala de 1km y por último la implementación y validación del sistema de monitorización de posibles derrumbes. En las siguientes secciones se detallarán cada uno de los procesos del proyecto.

En la primera parte y como primer acercamiento al problema, se realizó un análisis y extracción de *data* relevante de modelos de elevación digital de distintos terrenos afectados por deslizamientos, y con esto se obtuvo la información de la posición del terreno y la elevación de cada uno de los puntos de la superficie. Y luego este *DEM* obtenido también sirvió para obtener las curvas de nivel del área y la pendiente, aspecto, curvatura del plano y curvatura de perfil de cada uno de los puntos de la superficie del terreno utilizando la librería Richdem.

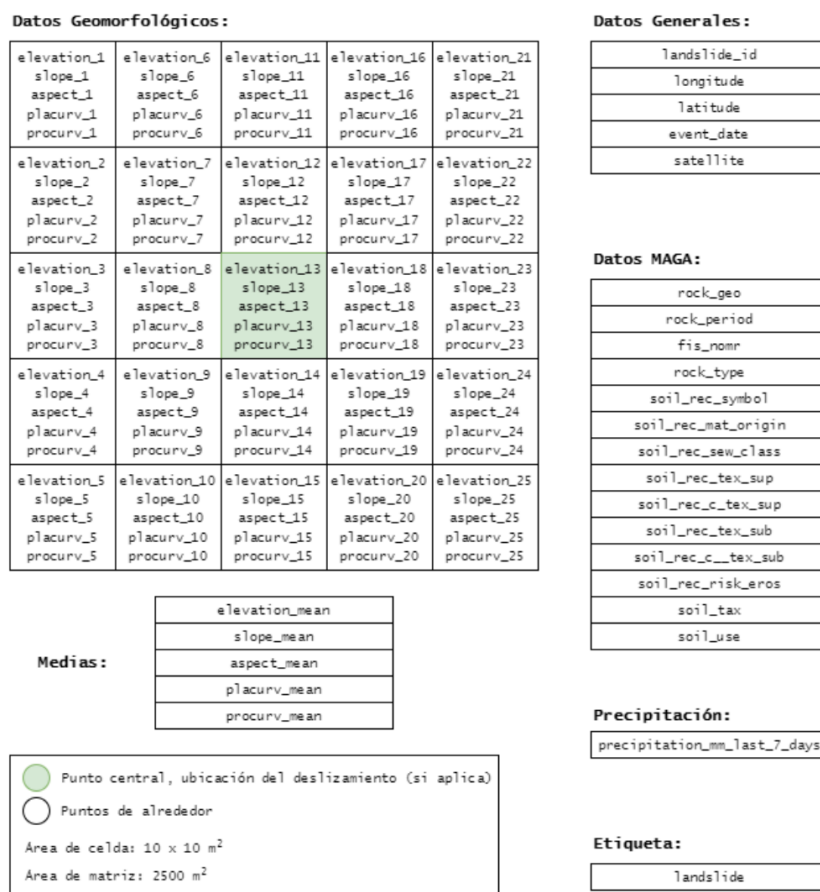
Figura 27: Diagrama de metodología general



Fuente: elaboración propia

El modelo de datos utilizado por el Modelo de susceptibilidad se encuentran en la Figura 28 y explicado de forma más específica en el Cuadro 20, del cual se hace referencia en varios de los procesos, especialmente del Proceso X que adjunta todos estos datos en un dataset. Los datos geomorfológicos de este modelo de datos corresponden a una matriz de datos la cual consta de 25 celdas de $10 \times 10 m^2$ como se puede ver en la Figura 29

Figura 28: Diagrama de modelo de datos



Fuente: elaboración propia

Cuadro 20: Modelo de datos creado para alimentar el modelo de susceptibilidad

<i>Feature</i>	Descripción	Valores	<i>Training</i>
landslide_id	Identificador ID del punto.	Numérico Int	
latitudo	Latitud de la posición a evaluar.	-90 a +90 EPSG:4386	
longitudo	Longitud de la posición a evaluar.	-180 a +180, EPSG:4386	
satellite	Satélite utilizado para obtener datos geomorfológicos.	Sentinel2, Landsat8 u Landsat7.	
event_date	Fecha del evento.	Fechas YY-MM-DD	
<i>precipitation_mm_last_7_days</i>	Precipitación acumulada de los últimos 7 días.	Medido en mm, valor numérico.	✓
<i>rock_period</i>	Periodo de roca de la posición.	Periodos de roca del mapa de geología (ver 8)	✓

<i>Feature</i>	Descripción	Valores	<i>Training</i>
<i>fis_nomr</i>	Fisiografía de la posición.	Fisiografía de Guatemala del mapa de Fisiografía (ver 13)	✓
<i>rock_type</i>	Tipo de roca de la posición.	Tipos de roca del mapa de geología (ver 8)	✓
<i>soil_rec_symbol</i>	Símbolo de reconocimiento del suelo de la posición.	Tipos de reconocimiento del mapa de reconocimiento de suelos Guatemala (ver 11)	✓
<i>soil_rec_mat_origin</i>	Material de origen del suelo.	Material de origen del suelo del mapa de reconocimiento de suelos Guatemala (ver 11)	✓
<i>soil_rec_sew_class</i>	Drenaje del suelo.	Drenaje del suelo del mapa de reconocimiento de suelos Guatemala (ver 11)	✓
<i>soil_rec_tex_sup</i>	Textura superior del suelo.	Textura superior del suelo del mapa de reconocimiento de suelos Guatemala (ver 11)	✓
<i>soil_rec_c_tex_sup</i>	Capacidad textura superior del suelo.	Capacidad textura superior del suelo del mapa de reconocimiento de suelos Guatemala (ver 11)	✓
<i>soil_rec_tex_sub</i>	Capacidad textura subterránea del suelo.	Capacidad textura subterránea del suelo del mapa de reconocimiento de suelos Guatemala (ver 11)	✓
<i>soil_rec_c_tex_sub</i>	Capacidad textura subterránea del suelo.	Capacidad textura subterránea del suelo del mapa de reconocimiento de suelos Guatemala (ver 11)	✓
<i>soil_rec_risk_eros</i>	Riesgo de erosión del suelo.	Riesgo de erosión del suelo del mapa de reconocimiento de suelos Guatemala (ver 11)	✓
<i>soil_tax</i>	Clasificación taxonómica del suelo	Clasificación taxonómica del suelo de mapa de Clasificación Taxonómica de Guatemala (ver 10)	✓
<i>soil_enguse</i>	Capacidad del uso del suelo.	Capacidad del uso del suelo del mapa de Capacidad de Uso de suelos Guatemala (ver 12)	✓
<i>elevation_n</i>	Elevación n de 25	Es la elevación n medida en m. (ver Figura 28)	✓
<i>slope_n</i>	Pendiente n de 25	Es la pendiente n medida en porcentaje. (ver Figura 28)	✓

<i>Feature</i>	Descripción	Valores	<i>Training</i>
aspect_n	Aspecto n de 25	Es el aspecto n siendo la dirección de la pendiente. (ver Figura 28)	✓
placurv_n	Curvatura n de 25	Es la curvatura del plano n. (ver Figura 28)	✓
procurv_n	Perfil n de 25	Es la curvatura del perfil n. (ver Figura 28)	✓
landslide	Etiqueta de entrenamiento para saber si existió un deslizamiento o no.	1 para deslizamiento y 0 para no deslizamiento.	✓

Fuente: elaboración propia

Figura 29: Diagrama de la matriz de datos geomorfológicos obtenidos del terreno

elevation_1 slope_1 aspect_1 placurv_1 procurv_1	elevation_6 slope_6 aspect_6 placurv_6 procurv_6	elevation_11 slope_11 aspect_11 placurv_11 procurv_11	elevation_16 slope_16 aspect_16 placurv_16 procurv_16	elevation_21 slope_21 aspect_21 placurv_21 procurv_21
elevation_2 slope_2 aspect_2 placurv_2 procurv_2	elevation_7 slope_7 aspect_7 placurv_7 procurv_7	elevation_12 slope_12 aspect_12 placurv_12 procurv_12	elevation_17 slope_17 aspect_17 placurv_17 procurv_17	elevation_22 slope_22 aspect_22 placurv_22 procurv_22
elevation_3 slope_3 aspect_3 placurv_3 procurv_3	elevation_8 slope_8 aspect_8 placurv_8 procurv_8	elevation_13 slope_13 aspect_13 placurv_13 procurv_13	elevation_18 slope_18 aspect_18 placurv_18 procurv_18	elevation_23 slope_23 aspect_23 placurv_23 procurv_23
elevation_4 slope_4 aspect_4 placurv_4 procurv_4	elevation_9 slope_9 aspect_9 placurv_9 procurv_9	elevation_14 slope_14 aspect_14 placurv_14 procurv_14	elevation_19 slope_19 aspect_19 placurv_19 procurv_19	elevation_24 slope_24 aspect_24 placurv_24 procurv_24
elevation_5 slope_5 aspect_5 placurv_5 procurv_5	elevation_10 slope_10 aspect_10 placurv_10 procurv_10	elevation_15 slope_15 aspect_15 placurv_15 procurv_15	elevation_20 slope_20 aspect_20 placurv_20 procurv_20	elevation_25 slope_25 aspect_25 placurv_25 procurv_25

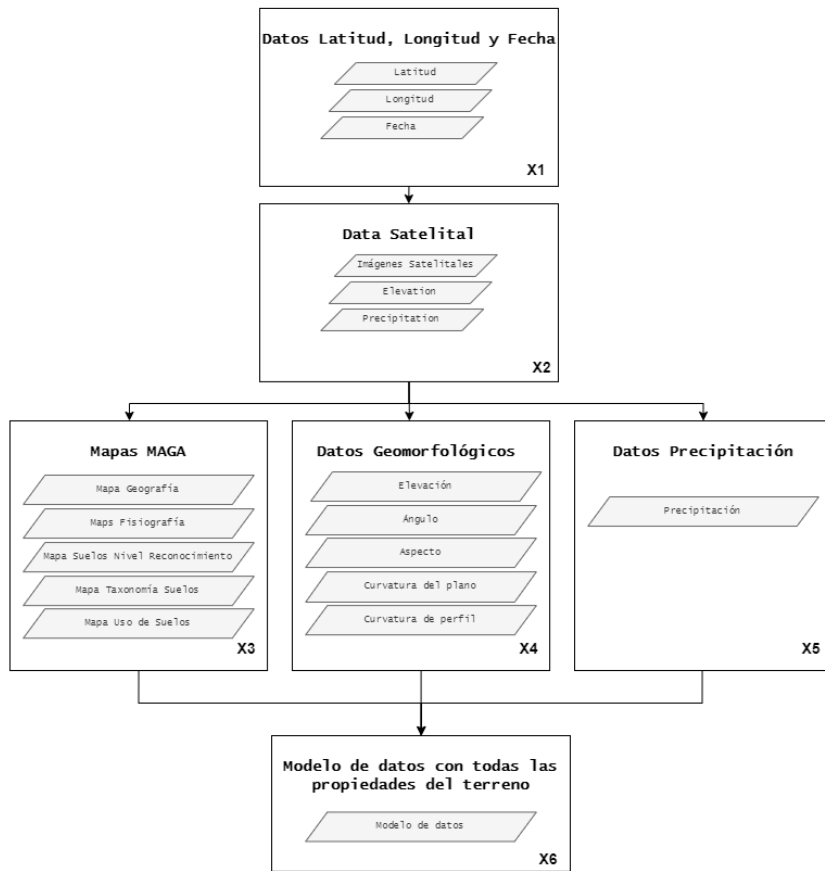
 Punto central, ubicación del deslizamiento (si aplica)
 Puntos de alrededor
Área de celda: 10 x 10 m ²
Área de matriz: 2500 m ²

Fuente: elaboración propia

5.2. Proceso X - Generación del modelo de datos

El proceso X es el encargado de generar el modelo de datos estándar del proyecto. Por esto mismo, los scripts realizados en este proceso son utilizados por muchos de los procesos del proyecto para generar la información necesaria para armar el modelo de datos. Por este motivo, este proceso no se agregó en la Figura 27, pero si en algunos de los procesos que se encuentren en él. El modelo de datos puede verse en el Cuadro 20.

Figura 30: Diagrama obtención de datos para modelo de susceptibilidad: proceso x



Fuente: elaboración propia

Para este proceso se utilizaron las siguientes herramientas tecnológicas, librerías y archivos de los cuadros 21, 22 y 23.

Cuadro 21: Herramientas tecnológicas utilizadas en el proceso x

Herramienta
Python 3.9.5

Fuente: elaboración propia

Cuadro 22: Librerías tecnológicas utilizadas en el proceso x

Librería	Descripción	Fuente
Numpy	Librería de Pandas para Python para manipulación de arrays y optimización de operaciones entre arrays.	Librería de Python
Pandas	Librería de Pandas para Python para la manipulación de <i>datasets</i> en <i>data-frames</i> .	Librería de Python
Shapely	Librería para la manipulación y análisis de <i>features</i> de un archivo <i>shape-file</i> .	Librería de Python
Richdem	Librería de Python que provee un conjunto de herramientas de análisis de modelos de elevación digital (<i>DEM</i>).	Librería de Python
Osgeo	Librería de Python que provee herramientas para manipular Geospatial <i>Data Abstraction Library</i> (GDAL).	Librería de Python
Datetime	Librería de Python para manejar variables de tipo date/time.	Librería de Python
ee	Librería de Python que permite obtener imágenes satelitales y realizar diferentes funciones con ellas.	Librería de Python
geetools	Librería que provee un conjunto de herramientas útiles para usar Google Earth Engine en Python.	Librería de Python

Fuente: elaboración propia

Cuadro 23: Archivos utilizados en el proceso x

Archivo	Descripción	Fuente
FISIOGRAFIAFINAL_GTM.shp	mapa de geografía de Guatemala	MAGA
GEOLOGIA_-GTM.shp	mapa de geología de Guatemala	MAGA
SSUELOS_GTM.shp	mapa de suelos nivel reconocimiento de Guatemala	MAGA
TAXONOMIA_OR-DENES_GTM.shp	mapa de taxonomía de suelos de Guatemala	MAGA
USDA_GTM.shp	mapa de uso de suelos de Guatemala	MAGA
<i>data-notebook.ipynb</i>	Notebook de Python que obtiene los datos de precipitación y datos geomorfológicos por cada punto en el mapa	Elaboración propia.
<i>data-notebook-maps.ipynb</i>	Notebook de Python que obtiene los datos de los mapas MAGA por cada punto en el mapa	Elaboración propia.
<i>data-notebook-merge.ipynb</i>	Notebook de Python que realiza un merge entre el csv generado por el archivo <i>data-notebook.ipynb</i> y el csv generado por el archivo <i>data-notebook-maps.ipynb</i>	Elaboración propia.
<i>landslides_satellite_images_get.ipynb</i>	Notebook de Python que obtiene las imágenes satelitales con su elevación.	Elaboración propia.
<i>landslides_precipitation_data_generation.ipynb</i>	Notebook de Python que obtiene las imágenes satelitales de precipitación y obtiene la precipitación acumulada de los últimos 7 días.	Elaboración propia.

Fuente: elaboración propia

Los pasos que se siguieron en este proceso se pueden observar en la Figura 30 y a continuación se presentan los pasos detalladamente:

X1 Datos latitud, longitud y fecha: Este proceso comienza con un archivo geojson, el cual posee la información de diferentes puntos en el mapa.

X2 Data satelital: En este paso se obtiene la *data* satelital de los puntos proporcionados con el geojson, utilizando la librería de earth engine y geetools de Python. Al principio del proyecto la información que se obtenía en este paso eran imágenes satelitales con el valor de elevación (*DEM*) de estas imágenes y la precipitación acumulada. Luego para la optimización de este proceso se obtenía directamente los valores de elevación del área analizada (sin descargar la imagen satelital) y el valor de la precipitación acumulada. Los pasos con mayor detalle son:

X2.1: Importar las herramientas de Google Earth Engine que permitirán la obten-

ción de imágenes satelitales (ver Cuadro 7). Además, es necesario cargar el dataset para el que se desea obtener toda esta información en el archivo *landslides_satellite_images_get.ipynb* y *landslides_precipitaion_data_generation.ipynb*.

X2.2: Por cada punto del dataset y sus respectivas coordenadas se obtiene la imagen satelital de dicho punto en la fecha específica, con la utilización del mejor satélite posible según sea dicha fecha y utilizando funciones de Masking de Nubes para reducir su interferencia en la imagen satelital. Luego de esto se proyecta con el *DEM* de la NASA la elevación de dicha imagen y se obtiene este valor que es necesario para los siguientes pasos. El código se presenta abajo:

```
# Get Satellite Images from Geojson
def get_satellite_available_images(geojson, days_to_check_for
    ↪ =180, bounds_meters=50, output_path=PATH_OUTPUT_ALL_IMAGES
    ↪ ):
    error_landslides_ids_all = []
    landslides_not_includes_in_all = []
    for \eng{feature} in geojson.\eng{features}:
        geometry = \eng{feature}.geometry
        coordinates = geometry.coordinates
        properties = \eng{feature}.properties
        landslide_id = properties[''landslide_id'']
        event_date = properties[''event_date'']
        landslide_prefix = ''' if properties[''landslide'']==''1''
            ↪ else ''non_''
        region = ee.Geometry.Point(coords=coordinates).buffer(
            ↪ bounds_meters).\eng{bounds}()
        end_date = datetime.strptime(event_date, '%Y-%m-%d') -
            ↪ timedelta(\eng{days}=1)
        init_date = datetime.strptime(event_date, '%Y-%m-%d') -
            ↪ timedelta(\eng{days}=days_to_check_for)
    try:
        if(init_date>sentinel_2_range[0]):
            collection_region = sentinel_2_collection.filterDate(
                ↪ str(init_date).\eng{split}('' ''')[0], str(
                ↪ end_date).\eng{split}('' ''')[0]).filterBounds(
                ↪ region).filter(ee.Filter.lte('
                ↪ CLOUDY_PIXEL_PERCENTAGE', 20)).map(
                ↪ sentinel_2_mask_clouds_function).sort('
                ↪ CLOUDY_PIXEL_PERCENTAGE', \eng{True})
            image_input = collection_region.first()

            #Calculate ndvi
            nir = image_input.select('B8')
            red = image_input.select('B4')
            blue = image_input.select('B2')
            ndvi = nir.subtract(red).divide(nir.add(red)).rename
                ↪ ('NDVI')
            green = image_input.select('B3')
```

```

gndvi = nir.subtract(green).divide(nir.add(green)).
    ↪ rename('GNDVI')

#Calculate Brightness
brightness = ((red.add(green).add(blue)).divide(3)).
    ↪ divide(1000)).rename('brightness')

#Calculate \eng{slope} \eng{and} \eng{elevation} from
    ↪ nasa dem \eng{and} adapt projection to input
    ↪ image
imageProjection = ndvi.projection()
\eng{slope}_dem = ee.\eng{Terrain}.\eng{slope}(
    ↪ nasa_dem).rename('\eng{slope}')
dem_projected = nasa_dem.reproject(imageProjection)
\eng{slope}_projected = \eng{slope}_dem.reproject(
    ↪ imageProjection)

#Merge the \eng{bands}
image = image_input.float().select('B4', 'B3', 'B2',
    ↪ 'B8').addBands(ndvi.float().select('NDVI'))
image = image.addBands(\eng{slope}_projected.float().
    ↪ select('\eng{slope}'))
image = image.addBands(dem_projected.float().select
    ↪ ('\eng{elevation}')).rename('\eng{elevation}')
image = image.addBands(gndvi.float().select('GNDVI'))
image = image.addBands(brightness.float().select('
    ↪ brightness'))

image_input_comp = gd.MaskedImage.from_id(str(image.
    ↪ id))
image_input_comp.ee_image = image
image_input_comp.download('{}{}
    ↪ _sentinel2_pre_landslide_test{}.tif'.format(
    ↪ output_path, landslide_prefix, landslide_id),
    ↪ region=region, scale=10,
                                crs='EPSG:4326',
                                overwrite=\eng{True}
    )
elif(init_date>landsat_8_range[0]):
    collection_region = landsat_8_collection.filterDate(
        ↪ str(init_date).\eng{split}(' '')[0], str(
        ↪ end_date).\eng{split}(' '')[0]).filterBounds(
        ↪ region).filter(ee.Filter.lte('CLOUD_COVER',
        ↪ 20)).map(landsat_mask_clouds_function).sort('
        ↪ CLOUD_COVER', \eng{True})
    image_input = collection_region.first()

```

```

#Calculate ndvi
nir = image_input.select('B5')
red = image_input.select('B4')
blue = image_input.select('B2')
ndvi = nir.subtract(red).divide(nir.add(red)).rename
    ↪ ('NDVI')
green = image_input.select('B3')
gndvi = nir.subtract(green).divide(nir.add(green)).
    ↪ rename('GNDVI')

#Calculate Brightness
brightness = ((red.add(green).add(blue)).divide(3)).
    ↪ divide(1000)).rename('brightness')

#Calculate \eng{slope} \eng{and} \eng{elevation} from
    ↪ nasa dem \eng{and} adapt projection to input
    ↪ image
imageProjection = ndvi.projection()
\eng{slope}_dem = ee.\eng{Terrain}.\eng{slope}(
    ↪ nasa_dem).rename('\eng{slope}')
dem_projected = nasa_dem.reproject(imageProjection)
\eng{slope}_projected = \eng{slope}_dem.reproject(
    ↪ imageProjection)

#Merge the \eng{bands}
image = image_input.float().select('B4', 'B3', 'B2',
    ↪ 'B5').addBands(ndvi.float().select('NDVI'))
image = image.addBands(\eng{slope}_projected.float().
    ↪ select('\eng{slope}'))
image = image.addBands(dem_projected.float().select
    ↪ ('\eng{elevation}').rename('\eng{elevation}'))
image = image.addBands(gndvi.float().select('GNDVI'))
image = image.addBands(brightness.float().select('
    ↪ brightness'))

image_input_comp = gd.MaskedImage.from_id(str(image.
    ↪ id))
image_input_comp.ee_image = image
image_input_comp.download('{{{
    ↪ _landsat8_pre_landsdlide_test_}}}.tif'.format(
    ↪ output_path, landslide_prefix, landslide_id),
    ↪ region=region, scale=10,
        crs='EPSG:4326',
        overwrite=\eng{True}
    )
elif (init_date>landsat_7_range[0]):
    collection_region = landsat_7_collection.filterDate(

```

```

    ↪ str(init_date).\eng{split}('' ')[0], str(
    ↪ end_date).\eng{split}('' ')[0]).filterBounds(
    ↪ region).filter(ee.Filter.lte('CLOUD_COVER',
    ↪ 20)).map(landsat_mask_clouds_function).sort('
    ↪ CLOUD_COVER', \eng{True})
image_input = collection_region.first()

#Calculate ndvi
nir = image_input.select('B4')
red = image_input.select('B3')
blue = image_input.select('B1')
ndvi = nir.subtract(red).divide(nir.add(red)).rename
    ↪ ('NDVI')
green = image_input.select('B2')
gndvi = nir.subtract(green).divide(nir.add(green)).
    ↪ rename('GNDVI')

#Calculate Brightness
brightness = ((red.add(green).add(blue)).divide(3)).
    ↪ divide(1000)).rename('brightness')

#Calculate \eng{slope} \eng{and} \eng{elevation} from
    ↪ nasa dem \eng{and} adapt projection to input
    ↪ image
imageProjection = ndvi.projection()
\eng{slope}_dem = ee.\eng{Terrain}.\eng{slope}(
    ↪ nasa_dem).rename('\eng{slope}')
dem_projected = nasa_dem.reproject(imageProjection)
\eng{slope}_projected = \eng{slope}_dem.reproject(
    ↪ imageProjection)

#Merge the \eng{bands}
image = image_input.float().select('B3', 'B2', 'B1',
    ↪ 'B4').addBands(ndvi.float().select('NDVI'))
image = image.addBands(\eng{slope}_projected.float().
    ↪ select('\eng{slope}'))
image = image.addBands(dem_projected.float().select
    ↪ ('\eng{elevation}').rename('\eng{elevation}'))
image = image.addBands(gndvi.float().select('GNDVI'))
image = image.addBands(brightness.float().select('
    ↪ brightness'))

image_input_comp = gd.MaskedImage.from_id(str(image.
    ↪ id))
image_input_comp.ee_image = image
image_input_comp.download('{{{
    ↪ _landsat7_pre_landslide_test_}}.tif'.format(
    ↪ output_path, landslide_prefix, landslide_id),

```

```

        ↪ region=region, scale=10,
                                                crs='EPSG:4326',
                                                overwrite=\eng{True}
    )
    else:
        landslides_not_includes_in_all.append(landslide_id)
except Exception as e:
    print(e)
    error_landslides_ids_all.append(landslide_id)
    landslides_not_includes_in_all.append(landslide_id)
    pass
print('Process \eng{of} getting images completed')
return landslides_not_includes_in_all, error_landslides_ids_all

```

X2.3: Por cada punto del dataset se obtiene también la precipitación acumulada de cada punto georeferenciado en la fecha específica, de los últimos 7 días, haciendo una suma de los valores diarios de cada hora en los últimos 7 días. El código se presenta abajo:

```

# \eng{Function} to get landslide \eng{precipitation} \eng{data} |
  ↪ \eng{of} geojson \eng{points} and append them to properties
def get_landslides_precipitation_data(geojson):
    error_landslides_ids_info_precipitation = []
    landslides_not_includes_in_info_precipitation = []
    count = 0
    for feature in geojson.features:
        count += 1
        geometry = feature.geometry
        coordinates = geometry.coordinates
        properties = feature.properties
        new_precipitation_properties = {}
        landslide_id = properties['landslide_id']
        event_date = properties['event_date']
        landslide_prefix = ''' if properties['landslide']==''1''
        ↪ else ''non_''
        region = ee.Geometry.Point(coords=coordinates)
        # \eng{Search} \eng{last} 7 \eng{days} \eng{of} \eng{
        ↪ precipitation}
        end_date = datetime.strptime(event_date, '%Y-%m-%d')
        init_date = datetime.strptime(event_date, '%Y-%m-%d') -
        ↪ timedelta(\eng{days}=7)
    try:
        #gsmmap_precipitation_collection_2014
        if(init_date>gsmmap_precipitation_collection_2014_range
        ↪ [0]):
            collection_region =
            ↪ gsmmap_precipitation_collection_2014.filterDate

```

```

        ↪ (str(init_date).split(' '')[0], str(end_date)
        ↪ .split(' '')[0]).filterBounds(region)
hourly_precip_rate_sum_dataset = collection_region.
        ↪ select('hourlyPrecipRate').sum()
pixel_info_hourly_precip_rate_sum =
        ↪ hourly_precip_rate_sum_dataset.reduceRegion(
            geometry=region,
            scale=10,
            reducer=ee.Reducer.mean()
        )

new_precipitacion_properties = {
    'hourly_precip_rate_sum':
        ↪ pixel_info_hourly_precip_rate_sum.get('
        ↪ hourlyPrecipRate').getInfo(),
}

#gsmap_precipitation_collection_2000
elif(init_date>gsmap_precipitation_collection_2000_range
        ↪ [0]):
    collection_region =
        ↪ gsmap_precipitation_collection_2000.filterDate
        ↪ (str(init_date).split(' '')[0], str(end_date)
        ↪ .split(' '')[0]).filterBounds(region)
hourly_precip_rate_sum_dataset = collection_region.
        ↪ select('hourlyPrecipRate').sum()
pixel_info_hourly_precip_rate_sum =
        ↪ hourly_precip_rate_sum_dataset.reduceRegion(
            geometry=region,
            scale=10,
            reducer=ee.Reducer.mean()
        )

new_precipitacion_properties = {
    'hourly_precip_rate_sum':
        ↪ pixel_info_hourly_precip_rate_sum.get('
        ↪ hourlyPrecipRate').getInfo(),
    # 'hourly_precip_rate_mean':
        ↪ pixel_info_hourly_precip_rate_mean.get('
        ↪ hourlyPrecipRate').getInfo()
}

else:
    landslides_not_includes_in_info_precipitation.
        ↪ append(landslide_id)

\eng{feature}.properties = dict(properties.items() |
        ↪ new_precipitacion_properties.items())

```



```

        print('{} / {} completed'.format(count, len(geojson.\len{
            ↪ features})))
    except Exception as e:
        print(e)
        error_landslides_ids_info_precipitation.append(
            ↪ landslide_id)
        landslides_not_includes_in_info_precipitation.append(
            ↪ landslide_id)
        pass
    return geojson, error_landslides_ids_info_precipitation,
        ↪ landslides_not_includes_in_info_precipitation

```

X2.4: Se guarda los datos con su respectivo `landslide_id` para luego poder enviar estos datos obtenidos al proceso X4 y X5 respectivamente.

X3 Mapas MAGA: Se obtuvo información relevante de cinco diferentes mapas, los cuales era un mapa de geografía, fisiología, suelos nivel reconocimiento, taxonomía del terreno y uso de suelos. Estos mapas estaban en formato de *shapefile* y cada *shapefile* contaba con una lista de *features* las cuales cada una relacionaba un área en Guatemala con información del terreno de dicha área. Se realizó el archivo *data-notebook-maps.ipynb* el cual realiza los siguientes pasos:

X3.1: Obtención de cada una de los áreas de cada mapa con su información relevante, utilizando la librería *shapely*

X3.2: Por cada punto del modelo en creación se recorren las áreas de los cinco mapas revisando si alguna de estas áreas contiene al punto (longitud, latitud).

X3.3: Si un área contiene al punto, se agrega la información de esta área al modelo en creación para dicho punto.

X3.4: Se guarda el modelo con la nueva información de los mapas en un archivo *shapefile*.

X4 Datos geomorfológicos: Se utilizó la elevación obtenida en el paso X2. A partir de esta elevación se obtuvo el ángulo, el aspecto, la curvatura del plano y la curvatura de perfil de cada uno de los puntos del modelo en creación. Esto se realizó en el archivo *data-notebook.ipynb* utilizando la función *terrain attributes* de la librería de Richdem de python. El archivo de python realiza los siguientes pasos:

X4.1: Se lee el directorio donde se almacenaron todas las imágenes satelitales utilizando la librería *osgeo*.

X4.2: Se obtienen todos los *bands* de la imagen (incluyendo el *band* de elevación).

X4.3: Itera sobre cada una de las imágenes satelitales obteniendo el id de la imagen (que será el `landslide_id` dentro del modelo de datos) y también obteniendo el *band* de elevación.

X4.4: Se obtiene los 25 puntos centrales del raster del *band* de elevación, los cuales cada celda tiene un área de $10 \times 10 m^2$ y todo el raster, o matriz tiene un área de $2500 m^2$. Estos 25 puntos se organizaron según la estructura de la Figura 29 y se agregaron al modelo de datos.

X4.5: Utilizando la elevación obtenida en el punto anterior, se calculó el ángulo, el aspecto, la curvatura del plano y la curvatura de perfil, utilizando la librería `richdem` (los cálculos que se realizan en esta librería se explican más adelante). Y luego se obtuvo los 25 puntos centrales del raster de cada uno de estos valores calculados (como se puede ver en la Figura 29) para así también agregarlos al modelo de datos.

X4.6: Por último se guarda el modelo con todos los datos calculados en un archivo `csv`.

X5 Datos de precipitación: Se utilizó el valor de precipitación obtenido en el paso X2. Este dato consistía en la precipitación acumulada de los últimos siete días en el área. Para la unión de esta información al principio para las primeras veces en que se realizó este proceso, se utilizó un script de python el cual relacionaba los Ids de ambos *datasets* (El modelo en creación y la información o dataset obtenido en el paso X2). Luego, para optimización del código, en el mismo archivo `data-notebook.ipynb` se iteraban los puntos obtenidos en el paso X2 y se calculaba tanto la información del paso X4 y X5, para luego crear un dataset con toda esta información por puntos.

X6 Modelo de datos con todas las propiedades del terreno: Se utilizó el archivo `data-notebook-merge.ipynb` para unir la información de cada punto, la cual consistía de la información de los mapas, los datos geomorfológicos y la precipitación acumulada de los últimos siete días. Para así crear el modelo de datos con todas estas propiedades del terreno por cada punto proporcionado al inicio del proceso. Por último, este modelo de datos se almacenaba en formato *shapefile* para asegurar la integridad de los datos.

Como ya se mencionó anteriormente, este proceso fue desarrollado al inicio del proyecto, pero se fue adaptando, mejorado y modificado según las necesidades del mismo, a medida que se avanzó en la realización del proyecto.

Por último, debe resaltarse que en el paso **X3**, se utilizan los mapas de Guatemala proporcionados por el Ministerio de Agricultura, Ganadería y Alimentación de Guatemala [MAGA] 22. El proceso de obtención de estos mapas de forma resumida es el siguiente (puede volverse un poco complejo por el tiempo de respuesta y el conocimiento de los datos disponibles):

1. Contactar con el ministerio de agricultura a través de correo electrónico: `uip@maga.gob.gt`, que es el departamento de información pública con la solicitud de información de los datos que se desean solicitar.
2. Se debe esperar la respuesta de este Departamento, para que luego se asigne la solicitud a un técnico del departamento de información geográfica, que posteriormente se comunicará con nosotros para especificar detalles de la información solicitada.
3. Luego de esta comunicación se debe esperar hasta 10 días hábiles para una respuesta, en la cual se enviarán los datos a través de un servicio de internet como Google Drive o algún otro, de los mapas solicitados. También, podrán decirnos que podemos llevar un dispositivo de almacenamiento al ministerio si así deseamos y los datos estarán listos preparados por el técnico.

4. Procedemos a descargar y revisar que sea la información que necesitamos e iterativamente volvemos a contactar si se desea mayor detalle u otra información.

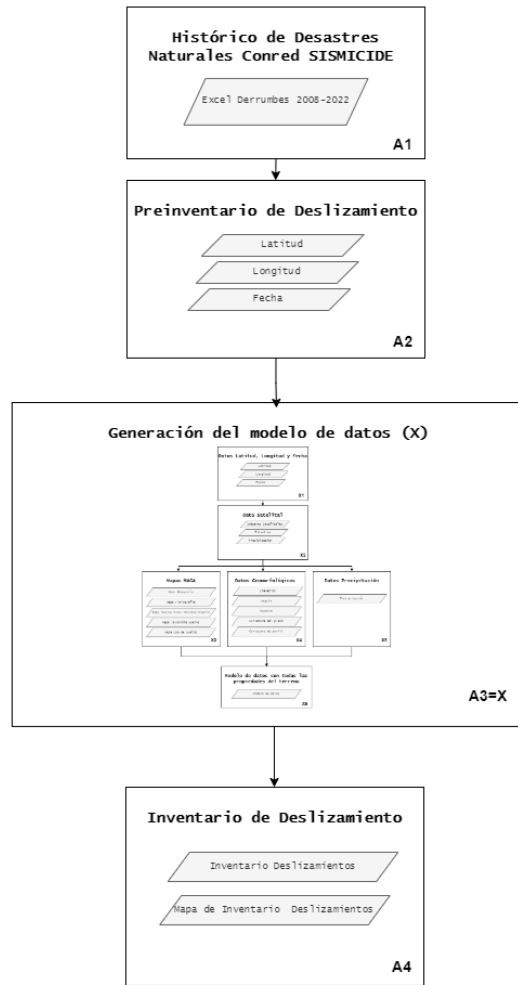
5.3. Proceso A - Generación de inventarios de deslizamientos en Guatemala

En el proceso A se utilizó el histórico de desastres naturales proporcionado por la Coordinación Nacional para la Reducción de Desastres de Guatemala [CONRED], el cual tenía información sobre derrumbes en Guatemala desde el 2008, hasta el 2022 [7]. Esta información fue de utilidad para crear un inventario de deslizamientos en Guatemala.

El historial de desastres es llevado por la CONRED, a través de Sistema de Manejo de Información en Caso de Emergencia o Desastre [SISMICEDE], el cual es un sistema de bases de datos que registra y almacena la información de eventos de desastres. La herramienta y su ingreso de desastres es de uso exclusivo de miembros del sistema Conred acreditados [7]. El proceso de obtención de estos datos de forma resumida es el siguiente (puede volverse un poco complejo por el tiempo de respuesta y el conocimiento de los datos disponibles):

1. Contactar con Conred a través de correo electrónico: infopublica.conred@gmail.com, que es el Departamento de Información Pública de la Conred con la solicitud de información de los datos que se desean solicitar.
2. Se debe esperar la respuesta de este Departamento, para que se asigne la solicitud, y que posteriormente se comuniquen con nosotros para especificar detalles de la información solicitada.
3. Luego de esta comunicación se debe esperar hasta 10 días hábiles para una respuesta, en la cual se enviarán los datos a través de un servicio de internet como Google Drive o algún otro, de los datos solicitados. También, podrán decirnos que podemos llevar un dispositivo de almacenamiento a SE-CONRED si así deseamos y los datos estarán listos preparados por un técnico.
4. Procedemos a descargar y revisar que sea la información que necesitamos e iterativamente volvemos a contactar si se desea mayor detalle u otra información.

Figura 31: Diagrama de generación del modelo de datos



Fuente: elaboración propia

Para este proceso se utilizaron las siguientes herramientas tecnológicas, librerías y archivos de los cuadros 24, 25 y 26.

Cuadro 24: Herramientas tecnológicas utilizadas en el proceso a

Herramienta
Python 3.9.5

Fuente: elaboración propia

Cuadro 25: Librerías tecnológicas utilizadas en el proceso a

Librería	Descripción	Fuente
Pandas	Librería de Pandas para Python para la manipulación de <i>datasets</i> en <i>data-frames</i> .	Librería de Python

Fuente: elaboración propia

Cuadro 26: Archivos utilizados en el proceso a

Archivo	Descripción	Fuente
derrumbes_2008-2022.xlsx	Archivo excel la información sobre derrumbes en Guatemala desde el 2008, hasta el 2022.	Elaboración propia
<i>landslides</i> .geojson	Archivo geojson con los puntos y la fecha en donde ocurrieron los deslizamientos en Guatemala (basados en el archivo derrumbes_2008-2022.xlsx)	Elaboración propia
<i>data_merged</i> .csv	Archivo csv con la información de cada punto del inventario de deslizamientos	Elaboración propia

Fuente: elaboración propia

Para ello, los pasos seguidos pueden observarse en la Figura 31 y constaron de los siguientes pasos:

- A1 Histórico de desastres naturales Conred SISMICIDE:** Este proceso comienza con el histórico de deslizamientos proporcionado por la Conred, el cual será útil para obtener los puntos de deslizamientos y se deben filtrar los registros únicamente por deslizamientos y derrumbes. Las fechas disponibles son desde el 2008, logrando generar obtener una cantidad de 4202 derrumbes entre el 2008 y el 2022.
- A2 Preinventario de deslizamiento:** Utilizando el histórico de deslizamientos se genera el archivo geojson *landslides*.geojson, el cual tiene la información de todos los puntos en donde se ha registrado un deslizamiento, así como su fecha.
- A3 Generación del modelo de datos:** En este paso se utilizaron los scripts realizados en el proceso de generación del modelo de datos (X). Estos scripts utilizaron los puntos obtenidos del preinventario de deslizamientos para agregarles la información necesaria para crear un modelo de datos en formato *shapefile*.
- A4 Inventario de deslizamientos:** Se obtuvo el inventario de deslizamientos utilizando toda la información del modelo de datos generada en el paso A3. Con este inventario de deslizamientos se logró obtener y visualizar por medio de *QGIS* los mapas de inventario de desplazamientos.

5.4. Proceso B - Generación dataset para entrenamiento de modelo de susceptibilidad a deslizamientos

Luego de haber obtenido, el inventario de deslizamientos es necesario generar los puntos fuera de zonas de riesgo de deslizamiento. Estos puntos, fuera de riesgo de deslizamiento, nos servirán para formar el dataset conjunto de todos los puntos necesarios para entrenar el modelo de *machine learning* de la parte C, ya que necesitará tanto puntos con etiquetas positivas (deslizamientos) y etiquetas negativas (no deslizamientos) para el aprendizaje. Cabe mencionar que esta parte fue mucha investigación de la herramienta *QGIS* y varias partes de sus funcionalidades, y se fue mejorando a través de comprender mejor el funcionamiento de la herramienta.

Cuadro 27: Herramientas tecnológicas utilizadas en el proceso b

Herramienta
<i>QGIS</i> Desktop 3.22.9
Python 3.9.5

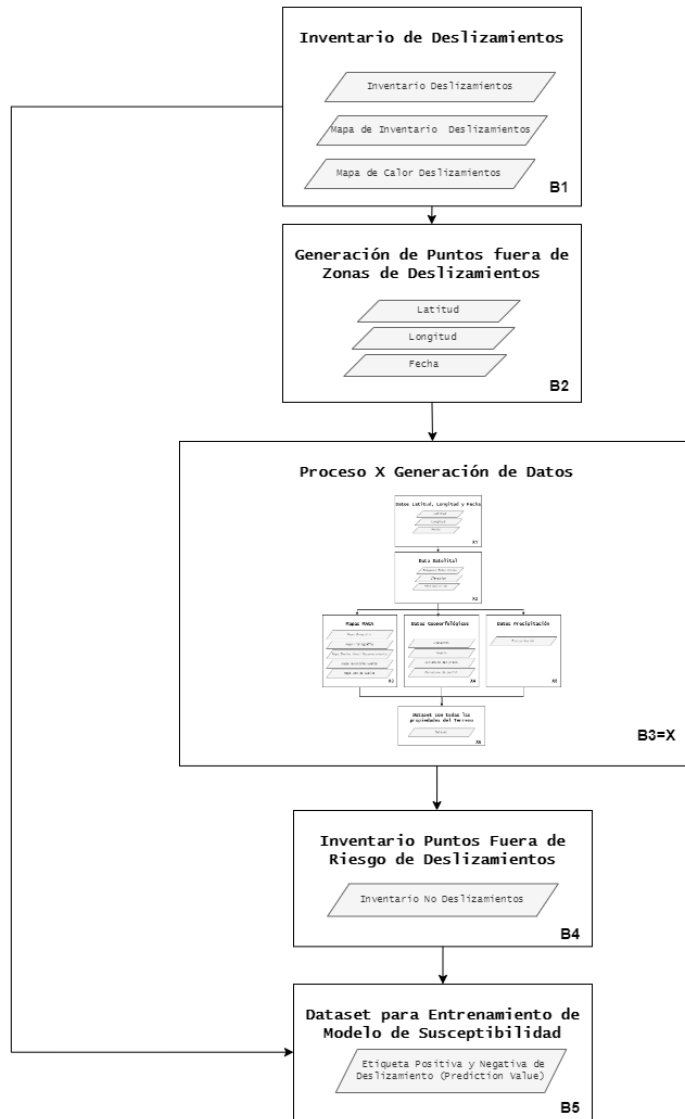
Fuente: elaboración propia

Cuadro 28: Librerías tecnológicas utilizadas en el proceso b

Librería	Descripción	Fuente
<i>Heatmap Plugin QGIS</i>	El complemento de heatmap utiliza <i>kernel density estimation</i> para crear un raster de densidad (mapa de calor) de una capa vectorial de puntos de entrada. La densidad se calcula en función del número de puntos en una ubicación, con un mayor número de puntos agrupados que dan como resultado valores más grandes.	<i>Plugin de QGIS</i>
<i>Raster Analysis QGIS</i>	El complemento de <i>Raster Analysis</i> de <i>QGIS</i> para realizar análisis de <i>Rasters</i> .	<i>Plugin de QGIS</i>
Google Earth Engine	<i>Plugin</i> desarrollado por Google para utilizar las herramientas de Google Earth Engine en <i>QGIS</i> .	<i>Plugin de QGIS</i>
Pandas	Librería de Pandas para Python para la manipulación de <i>datasets</i> en <i>dataframes</i> .	Librería de Python

Fuente: elaboración propia

Figura 32: Diagrama generación de dataset para entrenamiento de modelo de susceptibilidad a deslizamientos: proceso b



Fuente: elaboración propia

Para este proceso se utilizaron las siguientes herramientas tecnológicas, librerías y archivos de los cuadros 27, 28 y 29.

Cuadro 29: Archivos utilizados en el proceso b

Archivo	Descripción	Fuente
<i>non_landslides</i> .geojson	Este es el archivo de salida que contiene los puntos de no deslizamientos.	Elaboración propia a partir de <i>landslides</i> .geojson.
<i>landslides</i> .geojson	Este es el inventario de deslizamientos generado de la parte A.	Elaboración propia a partir de datos de Conred y MAGA.
cuerpos_agua_gtm.shp	Este es el mapa de cuerpos de agua de Guatemala.	Obtenido de Maga y Segeplan.
caminos_gtm.shp	Este es el mapa de caminos de Guatemala.	Obtenido de Maga y Segeplan.
departamentos_gtm.shp	Este es el mapa de departamentos de Guatemala.	Obtenido de Maga y Segeplan.
<i>data-notebook-merge</i> .ipynb	Notebook de python que realiza un merge entre el csv de puntos fuera de riesgo (<i>data_merged.csv</i>) y el csv de puntos con deslizamiento (<i>data_non_merged.csv</i>) para guardarlos en el archivo <i>data_all_merged.csv</i>	Elaboración propia.
<i>data_merged.csv</i>	Archivo csv generado en el proceso A con la información de cada punto del inventario de deslizamientos	Elaboración propia
<i>data_non_merged.csv</i>	Archivo csv con la información de cada punto fuera de riesgo de deslizamientos	Elaboración propia
<i>data_train.csv</i>	Archivo csv con el dataset para el entrenamiento del modelo de susceptibilidad a derrumbes	Elaboración propia

Fuente: elaboración propia

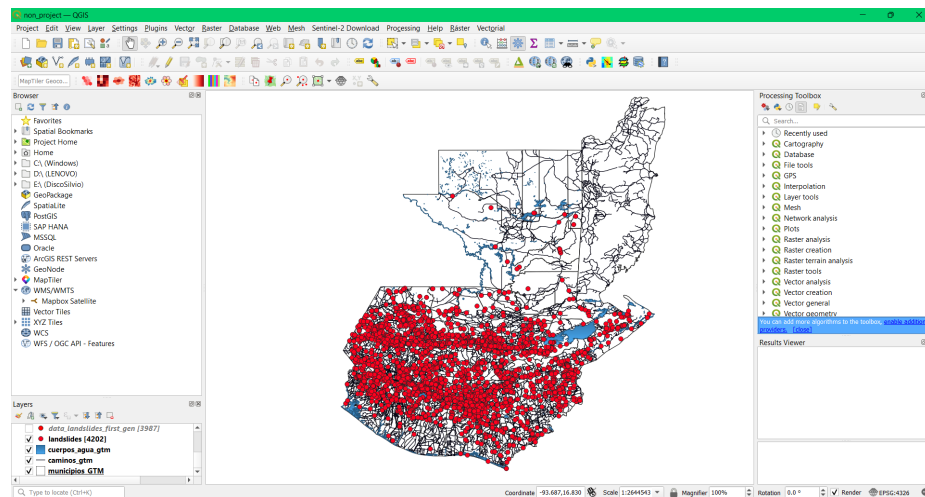
Los pasos seguidos pueden observarse en la Figura 32 y son explicados de la siguiente forma:

- B1 Inventario de deslizamientos:** Al obtener el inventario de deslizamientos se procede a realizar una exploración de los datos sumado a la investigación de las herramientas posibles para su tratamiento. En este caso, se aprende a utilizar los heatmaps y lo que representan en el inventario de deslizamientos. De igual forma, se aprende a manipular, categorizar y visualizar los datos en *QGIS* para poder seguir con el siguiente paso.
- B2 Generación de puntos fuera de zonas de riesgo de deslizamiento:** Durante este paso, se toma el inventario de deslizamiento que se obtuvo de la parte A, se genera un *heatmap* de deslizamientos y se utiliza la elevación de Google Earth Engine, para generar 8000 puntos de datos que correspondan a no deslizamientos, afirmando que estén fuera de 2km de las zonas más densas de deslizamientos y se encuentren en elevaciones planas a su alrededor. Se informa que este es el único proceso que se

ha realizado de forma manual (queriendo decir que no fue automatizado) ya que se realizó con la ayuda de *QGIS*, sus herramientas y basándose en la bibliografía existente sobre la generación de puntos de no deslizamientos. Esto se realiza por medios de los siguientes subpasos:

- B2.1:** Cargar a la herramienta de *QGIS*, el inventario de deslizamientos que se encuentra en formato geojson. Estos nos permitirá visualizar los puntos de deslizamiento en el mapa de Guatemala. De igual forma, es necesario cargar los mapas base de la República de Guatemala de departamentos, caminos viales y cuerpos de agua como se puede observar en la Figura 33. Cada uno de estos archivos puede ser cargado seleccionando **Layer > Add Layer > Add Vector Layer > (Agregar Archivo Vector) > Add**

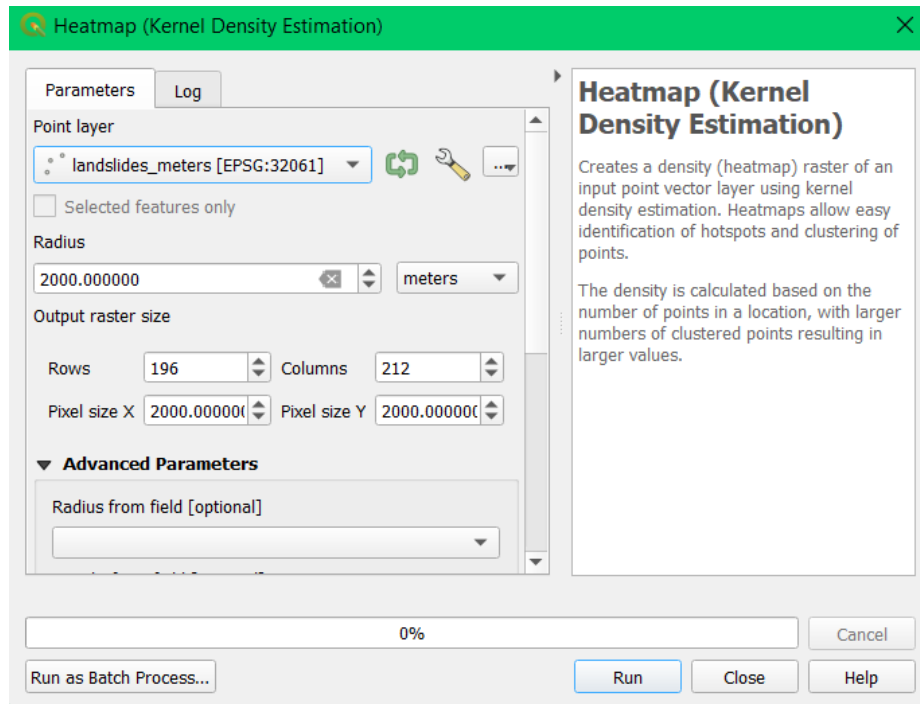
Figura 33: Inventario de deslizamientos en *QGIS*



Fuente: elaboración propia

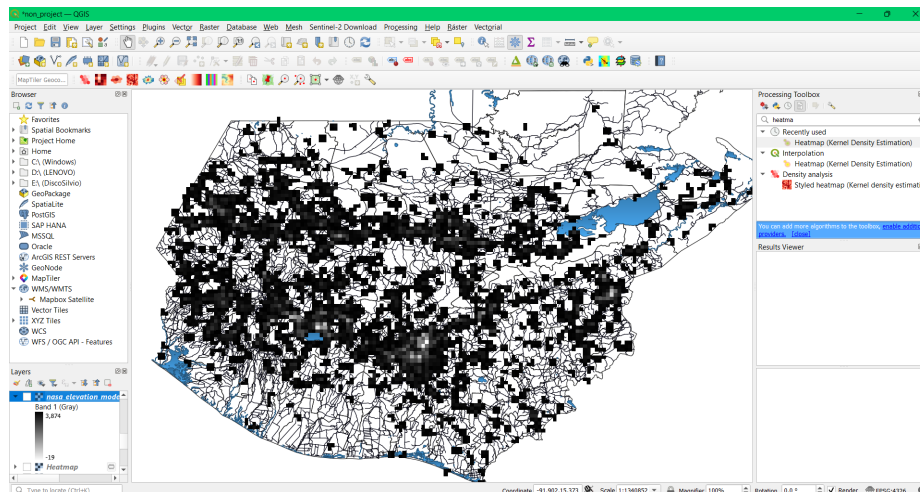
- B2.2:** Ahora se puede proceder a realizar un mapa de calor del inventario de deslizamientos con las herramientas disponibles en *QGIS*. Se deben seleccionar las siguientes opciones **Processing > Toolbox > Heatmap Kernel Density Estimation > Point Layer (Landslides) > Radius 2km > Run**. Esto permitirá generar raster de mapa de calor agrupando *landslides* por áreas de 2km. Puede observarse en la Figura 34, los parámetros utilizados para el *heatmap* y el raster generado en la Figura 35 donde se puede ver las zonas oscuras que son más propensas a derrumbes.

Figura 34: *Heatmap landslides raster params en QGIS*



Fuente: elaboración propia

Figura 35: *Heatmap landslides raster generado en qgis*



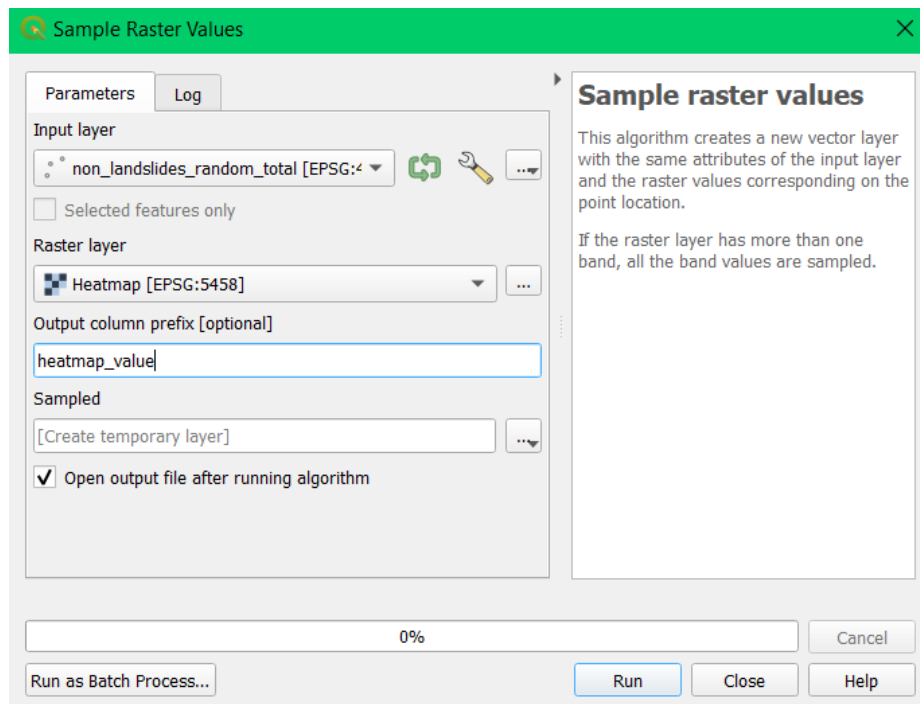
Fuente: elaboración propia

B2.3: Una vez obtenido el *raster* de mapa de calor de deslizamientos, se procede a agregar de catálogo de Google Earth Engine, la elevación y el *slope* para la región de Guatemala. A partir de esto, se pueden comenzar a generar los puntos aleatorios en el mapa de la República de Guatemala. Para ello se generan 10000 puntos aleatorios con los siguientes comandos (siendo estos puntos que se encuentran a

lo largo de las carreteras) **Vectorial** > **Herramientas de investigación** > **Heatmap Kernel Density Estimation** > **Random Points on Lines** > (**Number of points 10000** e **Input Layer caminos_gtm**) > **Run** y otros 10000 puntos aleatorios que se encuentren en los extents del mapa de Guatemala con las siguientes instrucciones **Vectorial** > **Herramientas de investigación** > **Heatmap Kernel Density Estimation** > **Random Points in Layer Bounds** > (**Number of points 10000** e **Input Layer mapa_gtm**) > **Run**. Luego de esto, se procede a unir ambos *layers* como los primeros 20000 *random points* generados.

B2.4: Luego ya con los *random points* se procede a agregar la elevación en el mapa así cómo el valor del *heatmap* con las siguientes instrucciones **Processing** > **Toolbox** > **Raster Analysis** > **Sample raster values** > (**Raster Layer heatmap_landslides** e **Input Layer random_points**) > **Run**. Este se observa en la Figura 36. Esto agrega una columna a nuestro *layer* de *random_points* con el valor del raster *heatmap* siendo los valores de nuestro interés cuando este se encuentra valores cercanos a 0 y en lugares planos, por ello se hace el mismo proceso con el raster de elevación.

Figura 36: Sample raster values del landslide heatmap en QGIS



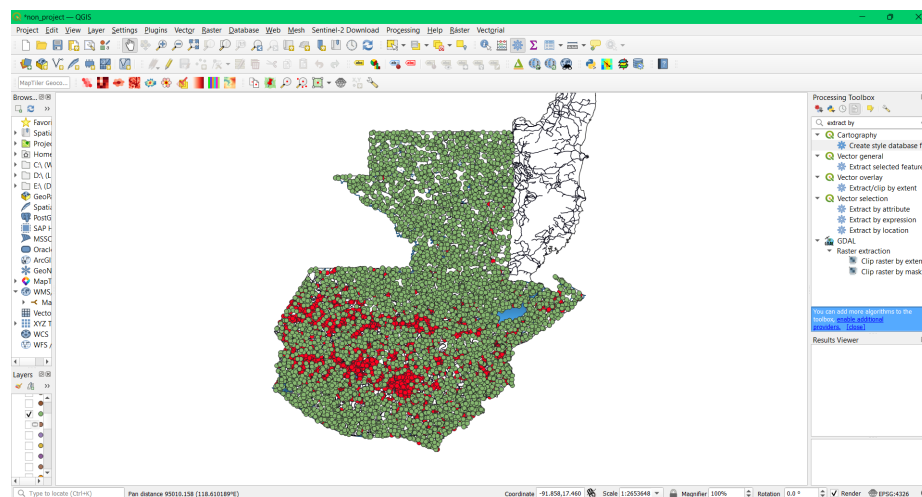
Fuente: elaboración propia

B2.5: Ya con los puntos teniendo los valores de valor de cercanía a áreas de deslizamiento (*heatmap*) y elevación. Se procede a realizar una extracción de estos puntos, de los que cuentan con las características que necesitamos. Para ello, procedemos a obtener los *random points* que se encuentran en zonas de fuera de deslizamiento (valor en el *heatmap* menor a 0.5) y se encuentren en zonas planas (interpretando la elevación de dichos puntos). Esto permite tener valores fuera

de zona de deslizamiento y dando una mayor a importancia a zonas planas que sabemos no tendrán deslizamientos. Esto se hace con las siguientes instrucciones: **Processing > Toolbox > Vector Extraction > Extract By Expression > (Input Layer Random Points y Expression es *heatmap* menor a 0.5 con elevaciones planas)> Run**. Esto nos permite obtener nuestro dataset final de puntos de no deslizamiento con un total de 8000 puntos. En principio, el doble de la cantidad de deslizamientos.

B2.6: Por último, ya se puede exportar la capa en formato geojson para realizar los siguientes pasos y posteriormente, unirse al dataset de los *landslides* para que este se complete con los puntos positivos y negativos. Este dataset puede verse en la Figura [37](#)

Figura 37: Dataset con *landslides* y *non landslides* en QGIS



Fuente: elaboración propia

B3 Proceso X a los datos generados En este paso se utilizaron los scripts realizados en el proceso de generación del modelo de datos (X). Estos scripts utilizaron los puntos obtenidos de no deslizamiento para agregarles la información necesaria para crear un modelo de datos en formato *shapefile*.

B4 Inventario puntos fuera de riesgo de deslizamiento Ahora se tiene un inventario de los puntos de no deslizamientos con todos los datos pertinentes del proceso X.

B5 Dataset para entrenamiento de modelo de susceptibilidad En este paso, se toma el inventario de deslizamiento del proceso A y el inventario de no deslizamientos del proceso B, para ser unidos en un solo archivo a través de Python y Pandas, para obtener el dataset de entrenamiento para el modelo de susceptibilidad.

5.5. Proceso C - Entrenamiento modelos de *machine learning* con dataset

Durante este proceso, se procede a entrenar varios modelos de *machine learning*, específicamente de aprendizaje supervisado en una clasificación binaria, evaluar sus resultados y posteriormente, guardar los modelos con los mejores resultados. Se realiza evaluación de los modelos a través de métricas tales como *ROCAUC*, *sensitivity*, *recall*, *precision* y *accuracy*. Este proceso es iterativo haciendo mejoras a los parámetros y *features*, logrando finalmente un modelo de susceptibilidad de diferentes algoritmos para posteriormente compararlos y evaluar sus resultados. Los algoritmos finalmente utilizados fueron basados en árboles: *Decision tree*, *Random forest* y *XGBoost* y basados en probabilística y estadística: *Logistic regression* y *Support vector machine*.

Cuadro 30: Herramientas tecnológicas utilizadas en el proceso c

Herramienta
Python 3.9.5

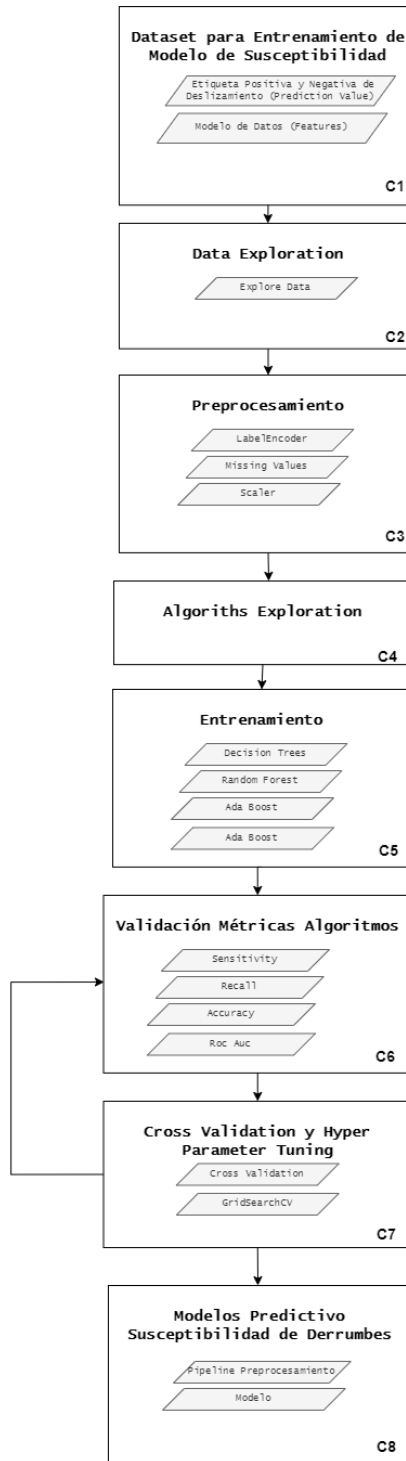
Fuente: elaboración propia

Cuadro 31: Librerías tecnológicas utilizadas en el proceso c

Librería	Descripción	Fuente
Pandas	Librería de Pandas para Python para la manipulación de <i>datasets</i> en <i>dataframes</i> .	Librería de Python
Sklearn	Librería de desarrollo por Scikit Learn para Python que contiene diferentes modelos de <i>machine learning</i> , al igual que las funciones para evaluar sus resultados.	Librería de Python
<i>XGBoost</i>	Librería de desarrollo por <i>XGBoost</i> para utilizar el algoritmo <i>XGBoost</i> en Python	Librería de Python

Fuente: elaboración propia

Figura 38: Diagrama generación de dataset para entrenamiento de modelo de susceptibilidad a deslizamientos: proceso c



Fuente: elaboración propia

Para este proceso se utilizaron las siguientes herramientas tecnológicas, librerías y archivos de los cuadros 30, 31 y 32.

Cuadro 32: Archivos utilizados en el proceso c

Archivo	Descripción	Fuente
<i>data_train.csv</i>	Este el archivo contiene el dataset de entrenamiento de derrumbes que fueron utilizados por los modelos para entrenar.	Elaboración propia a partir de datos de Conred y MAGA
<i>landslides_train_with_our_db.ipynb</i>	Este es un Notebook de Python que explora, preprocesa y entrena con varios modelos a partir del dataset.	Elaboración propia
<i>landslides_train_with_our_db_final_models.ipynb</i>	Este es un Notebook de Python que tunea los mejores parámetros para entrenar el modelo y obtiene los resultados pertinentes.	Elaboración propia

Fuente: elaboración propia

Los pasos seguidos pueden observarse en la Figura 38 y son explicados de la siguiente forma:

- C1 Dataset para entrenamiento de modelo de susceptibilidad:** En este paso se carga el dataset de entrenamiento contenido en el archivo csv (ver Fig 32) hacia un dataframe de Pandas que permita contener este dataset en nuestro ambiente en el archivo *landslides_train_with_our_db.ipynb*.
- C2 Data exploration:** En este paso se exploran los datos del dataset que hemos venido construyendo, viendo todos los posibles atributos, los tipos de datos y construyendo visualizando importantes.
- C3 Preprocesamiento:** Para esta paso, se realiza el preprocesamiento necesario por cada una de las *features*, por ejemplo todas las variables categóricas necesitan pasarse por un LabelEncoder, para ser convertidas a un dato numérico, dependiendo del algoritmo se necesita escalar y normalizar los datos y también se necesita tomar en cuenta las columnas con valores faltantes.
- C4 Algorithms exploration:** Aquí se exploran diferentes algoritmos, se investigan sus propiedades y posteriormente, se procede a entrenar y evaluar rápidamente estos algoritmos con ayuda de sklearn. Los algoritmos aquí contemplados son varios como *DecisionTreeClassifier*, *RandomForestClassifier*, *LogisticRegression*, *AdaBoostClassifier*, *XGBoostClassifier*, *SVMClassifier*, *NeuralNetworks*, *BaggingClassifier* y *NaiveBayesClassifier*.
- C5 Entrenamiento:** Se procede a realizar una evaluación únicamente de los algoritmos basados en árboles de decisión, por lo que los seleccionados son *DecisionTreeClassifier*, *RandomForestClassifier*, *XGBoostClassifier*, *LogisticRegression* y *SupportVectorMachine*.

- C6 Validación de métricas de algoritmos:** Se validan las métricas resultantes de los algoritmos de la parte C5.
- C6 Validación de métricas de algoritmos:** Se validan las métricas resultantes de los algoritmos de la parte C5 con métricas como *accuracy*, *sensitivity*, *recall*, *confussion matrix* y *roc auc*.
- C7 Cross validation y Hyperparameter tuning:** Con técnicas de *cross validation* y *Hyperparameter tuning* se procede a validar las mejores métricas y encontrar el mejor modelo para los algoritmos del paso C5.
- C8 Modelo predictivo susceptibilidad de derrumbes:** Se entrena finalmente con todo el dataset a los modelos con los mejores parámetros encontrados para ser guardados y poder ser utilizados en el Proceso E de evaluación de los modelos de susceptibilidad a nivel de Guatemala.

5.6. Proceso D - Generación puntos de Guatemala a 1km de distancia e integración de datos ligados al modelo

Durante este proceso, se generaron puntos en toda Guatemala y a 1 km de distancia entre cada punto. Se obtuvo también por cada uno de estos puntos, las mismas variables del modelo de datos del proceso B, pero lógicamente con mayor cantidad de puntos. De esta forma se obtuvo el dataset de toda Guatemala para poder evaluar el modelo de susceptibilidad.

Para este proceso se utilizó las siguientes herramientas tecnológicas, librerías y archivos de los cuadros [33](#), [34](#) y [35](#).

Cuadro 33: Herramientas tecnológicas utilizadas en el proceso d

Herramienta
Python 3.9.5

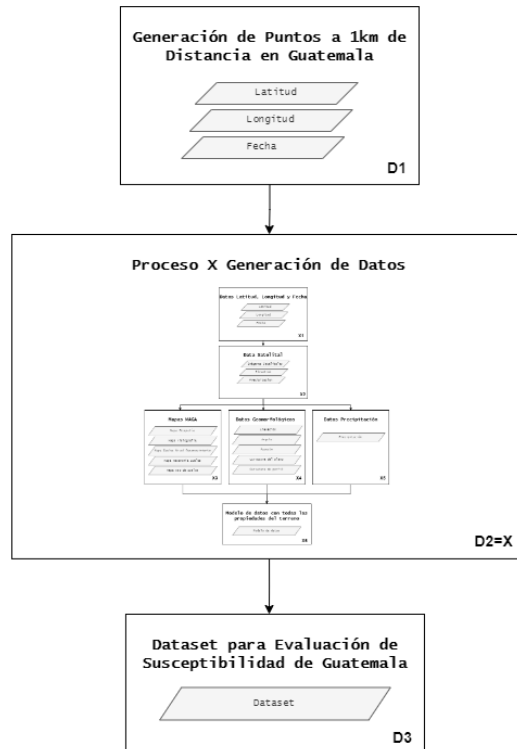
Fuente: elaboración propia

Cuadro 34: Librerías tecnológicas utilizadas en el proceso d

Librería	Descripción	Fuente
Pandas	Librería de Pandas para Python para la manipulación de <i>datasets</i> en <i>dataframes</i> .	Librería de Python
Shapely	Librería para la manipulación y análisis de <i>features</i> de un archivo <i>shapefile</i> .	Librería de Python

Fuente: elaboración propia

Figura 39: Diagrama generación de puntos de Guatemala a 1km de distancia e integración de datos ligados al modelo: proceso D



Fuente: elaboración propia

Cuadro 35: Archivos utilizados en el proceso d

Archivo	Descripción	Fuente
gadm41_GTM_0.shp	mapa en formato <i>shapefile</i> , con solo una <i>feature</i> de todo el terreno de Guatemala.	MAGA
<i>data-notebook-complete_map.ipynb</i>	Este es un Notebook de Python que obtiene puntos en toda Guatemala a un km de distancia entre cada punto.	Elaboración propia
<i>data_all_points_gt_map.csv</i>	Este el archivo contiene el dataset de los puntos de toda Guatemala a una distancia de 1 km entre sí.	Elaboración propia
<i>data_all_gt_map.csv</i>	Este el archivo contiene el dataset con el modelo de datos de los puntos de toda Guatemala a una distancia de 1 km entre sí.	Elaboración propia

Fuente: elaboración propia

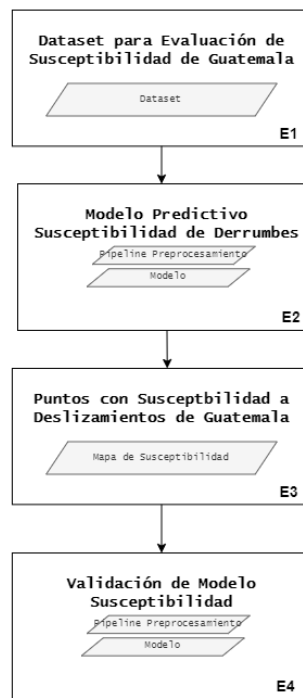
Los pasos seguidos pueden observarse en la Figura 39 y son explicados de la siguiente forma:

- D1 Generación de puntos a 1 km de distancia en Guatemala:** En este paso se realizó el archivo *data-notebook-complete_map.ipynb* el cual generaba 106,123 puntos en toda Guatemala a 1 km de distancia. El archivo realizaba los siguientes pasos:
- D1.1:** Obtención del *feature* de toda Guatemala del archivo *gadm41_GTM_0.shp*, utilizando la librería *shapely*,
 - D1.2:** Obtiene los *bounds* del *feature* del mapa de toda Guatemala.
 - D1.3:** Iteración sobre los *bounds* del *feature* de izquierda a derecha y de abajo hacia arriba, avanzando en cada iteración 1 km de distancia hacia la derecha, y luego al llegar al límite del mapa de la derecha se vuelve a comenzar desde la izquierda del mapa, pero ahora 1 km más hacia arriba, hasta llegar hasta la esquina superior izquierda del mapa.
 - D1.4:** En cada iteración, si el punto pertenece al mapa, este se guarda dentro de un dataframe con la fecha actual.
 - D1.5:** se guarda el dataframe generado en un archivo provisional con los puntos generados *data_all_points_gt_map.csv*.
- D2 Proceso X generación de datos:** En este paso se utilizaron los scripts realizados en el proceso de generación del modelo de datos (X). Estos scripts utilizaron los puntos de toda Guatemala para agregarles la información necesaria para crear un modelo de datos estándar del proyecto.
- D3 Dataset para evaluación de susceptibilidad de Guatemala:** Por último se obtiene el modelo de datos del paso D2 que utiliza el proceso X y se guarda en el archivo *data_all_gt_map.csv* que contiene toda la información del modelo de datos de los 106,123 puntos en toda Guatemala.

5.7. Proceso E - Evaluación de susceptibilidad de derrumbes en toda Guatemala escala 1km

Durante este proceso, se procede con los modelos entrenados del proceso C, a realizar la evaluación de susceptibilidad en el dataset de puntos del proceso D, que es una evaluación de susceptibilidad de derrumbes en toda Guatemala a una escala de 1km y asimismo, realizar la validación de cada uno de estos modelos (como las primeras visualizaciones de los mapas de susceptibilidad y categorizaciones de susceptibilidad).

Figura 40: Diagrama de evaluación de susceptibilidad de derrumbes en toda Guatemala escala 1km: proceso E



Fuente: elaboración propia

Para este proceso se utilizaron las siguientes herramientas tecnológicas, librerías y archivos de los cuadros 36, 37 y 38.

Cuadro 36: Herramientas tecnológicas utilizadas en el proceso e

Herramienta
QGIS 3.22
Python 3.9.5

Fuente: elaboración propia

Cuadro 37: Librerías tecnológicas utilizadas en el proceso e

Librería	Descripción	Fuente
Pandas	Librería de Pandas para Python para la manipulación de <i>datasets</i> en <i>dataframes</i> .	Librería de Python
Sklearn	Librería de desarrollado por Scikit Learn para Python que contiene diferentes modelos de <i>machine learning</i> , al igual que las funciones para evaluar sus resultados.	Librería de Python
<i>XGBoost</i>	Librería de desarrollado por <i>XGBoost</i> para utilizar el algoritmo <i>XGBoost</i> en Python	Librería de Python

Fuente: elaboración propia

Cuadro 38: Archivos utilizados en el proceso e

Archivo	Descripción	Fuente
<i>data_merged_all_gt_map_final.csv</i>	Este el archivo contiene el dataset de todos los puntos para ser utilizado para evaluar el modelo de susceptibilidad.	Elaboración propia a partir de datos de Conred y MAGA
<i>landslides_test_guatemala_points.ipynb</i>	Este es un Notebook de Python que realiza el <i>test</i> de todos los puntos de Guatemala para generar un geojson de la susceptibilidad.	Elaboración propia
<i>data_gt_map_geojson_final.geojson</i>	Este es el archivo geojson final con los datos de susceptibilidad de todos los modelos.	Elaboración propia

Fuente: elaboración propia

Los pasos seguidos pueden observarse en la Figura 40 y son explicados de la siguiente forma:

- E1 Dataset para evaluación de susceptibilidad:** En este paso se carga el dataset de evaluación contenido en el archivo csv (ver Fig. 32) hacia un dataframe de Pandas que permita contener este dataset en nuestro ambiente en el archivo *landslides_test_guatemala_points.ipynb*.
- E2 Modelo predictivo susceptibilidad de derrumbes:** Por cada uno de estos puntos, y por cada uno de los modelos producidos en el proceso C, se aplica la susceptibilidad y se genera este valor.
- E3 Puntos de susceptibilidad a deslizamientos de Guatemala:** Con los valores generados se adjuntan al dataset de la evaluación de susceptibilidad y se convierten

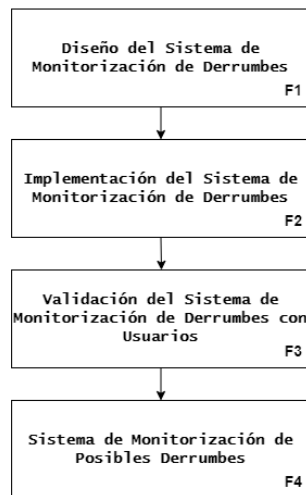
en un geojson, con estos valores de susceptibilidad predichos por cada modelo en el archivo `data_gt_map_geojson_final.geojson`.

E4 Validación de modelos de susceptibilidad: Con los valores generados de susceptibilidad y los datos de entrenamiento, se puede validar las métricas receiving operator characteristics, *Relative Error*, *relative landslide density index* y *density index* del Cuadro 5 para la evaluación de cada uno de estos modelos. Esto puede hacerse por un programa de Python, que haga la interpolación de los datos de susceptibilidad generados y los datos de entrenamiento, o ya bien por medio de *QGIS* tomando los valores y realizando los cálculos pertinentes que se establecen para cada métrica en el Cuadro 5.

5.8. Proceso F: Implementación y validación del sistema de monitorización de posibles derrumbes

Luego de aprobar y validar los mejores modelos de susceptibilidad, se pasa al proceso F, donde a partir del Diseño del Sistema de la Figura 61, se busca implementar y validar el Sistema de Monitorización de Derrumbes con los pasos de la Figura 41. Esta validación final se realizó mostrando los resultados a Ingenieros Civiles expertos en el tema.

Figura 41: Diagrama de implementación y validación del sistema de monitorización de posibles derrumbes: proceso f



Fuente: elaboración propia

5.9. Proceso Z - Entrenamiento de modelo de segmentación semántica

Luego de generar el mapa de susceptibilidad de derrumbes, se quería identificar semánticamente los terrenos monitorizados, y así poder identificar las diferentes partes de la imagen como bosques, terrenos de agricultura, carreteras y edificaciones. Para esto se utilizó un conjunto de datos para segmentación semántica, el cual se obtuvo en [41]. Este contiene 5987 imágenes HSR con, 166768 objetos anotados de tres ciudades diferentes y abarcan dos dominios (urbano y rural). Con este *set* de datos se entrenaron varios modelos para segmentación semántica de imágenes utilizando la librería de pytorch.

Se utilizó la librería de python *albumentations* para la variabilidad de los datos para entrenamiento y que los modelos entrenados tuvieran buenos resultados para imágenes aéreas y ortomosaicos, no importando la escala de estos mapas. Con esta se aplicaron transformaciones a las imágenes del entrenamiento y de validación. Estas transformaciones hicieron que las imágenes variaran en: tamaño, escala, rotación, brillo y sombras.

Los modelos utilizados para este proceso fueron: U-Net, *Feature Pyramid Network* (FPN), *Pyramid Scene Parsing Network* (PSPNet), *Path Aggregation Network* (PANet) y *DeepLab V3+*. Cada uno de estos modelos fueron evaluados por la métrica IoU de cada una de las clases que identificaba cada modelo, las cuales fueron: *no-data*, *background*, *building*, *road*, *forest* y *agriculture*. Y también se evaluó la media de la métrica IoU de cada clase (mIoU), la cual también sirvió para evaluar cada uno de los modelos.

Para este proceso se utilizaron las siguientes herramientas tecnológicas, librerías y archivos de los cuadros [39], [40] y [41].

Cuadro 39: Herramientas tecnológicas utilizadas en el proceso z

Herramienta
Python 3.9.5

Fuente: elaboración propia

Cuadro 40: Librerías tecnológicas utilizadas en el proceso z

Librería	Descripción	Fuente
Numpy	Librería de Pandas para Python para manipulación de arrays y optimización de operaciones entre arrays.	Librería de Python
Pandas	Librería de Pandas para Python para la manipulación de <i>datasets</i> en <i>data-frames</i> .	Librería de Python
Sklearn	Librería de desarrollo por Scikit Learn para Python que contiene diferentes modelos de <i>machine learning</i> , al igual que las funciones para evaluar sus resultados.	Librería de Python
PyTorch	Librería que facilita y optimiza el manejo de modelos de <i>machine learning</i>	Librería de Python
PIL	Librería de manejo y procesamiento de imágenes de python.	Librería de Python
OpenCV	Librería open source de <i>machine learning</i> y visión por computador de python.	Librería de Python
albumentations	Es una librería enfocada en visión por computador para manipular, distorsionar y modificar imágenes rápidamente y enfocada en el entrenamiento de redes neuronales.	Librería de Python
torchgeo	Es una librería de python que provee <i>datasets</i> y modelos preentrenados, para que estos puedan ser utilizados de una manera más sencilla.	Librería de Python
segmentation_models_pytorch	Es una librería de python con redes neuronales para segmentación semántica de imágenes basada en PyTorch.	Librería de Python

Fuente: elaboración propia

Cuadro 41: Archivos utilizados en el proceso z

Archivo	Descripción	Fuente
segmentation-notebook-Unet.ipynb	Este es un Notebook de Python que entrena un modelo Unet para segmentación semántica de imágenes.	Elaboración propia
segmentation-notebook-FPN.ipynb	Este es un Notebook de Python que entrena un modelo FPN para segmentación semántica de imágenes.	Elaboración propia
segmentation-notebook-PSPNet.ipynb	Este es un Notebook de Python que entrena un modelo <i>PSPNet</i> para segmentación semántica de imágenes.	Elaboración propia
segmentation-notebook-PAN.ipynb	Este es un Notebook de Python que entrena un modelo PAN para segmentación semántica de imágenes.	Elaboración propia
segmentation-notebook-DeepLabV3Plus.ipynb	Este es un Notebook de Python que entrena un modelo <i>DeepLab V3+</i> para segmentación semántica de imágenes.	Elaboración propia

Fuente: elaboración propia

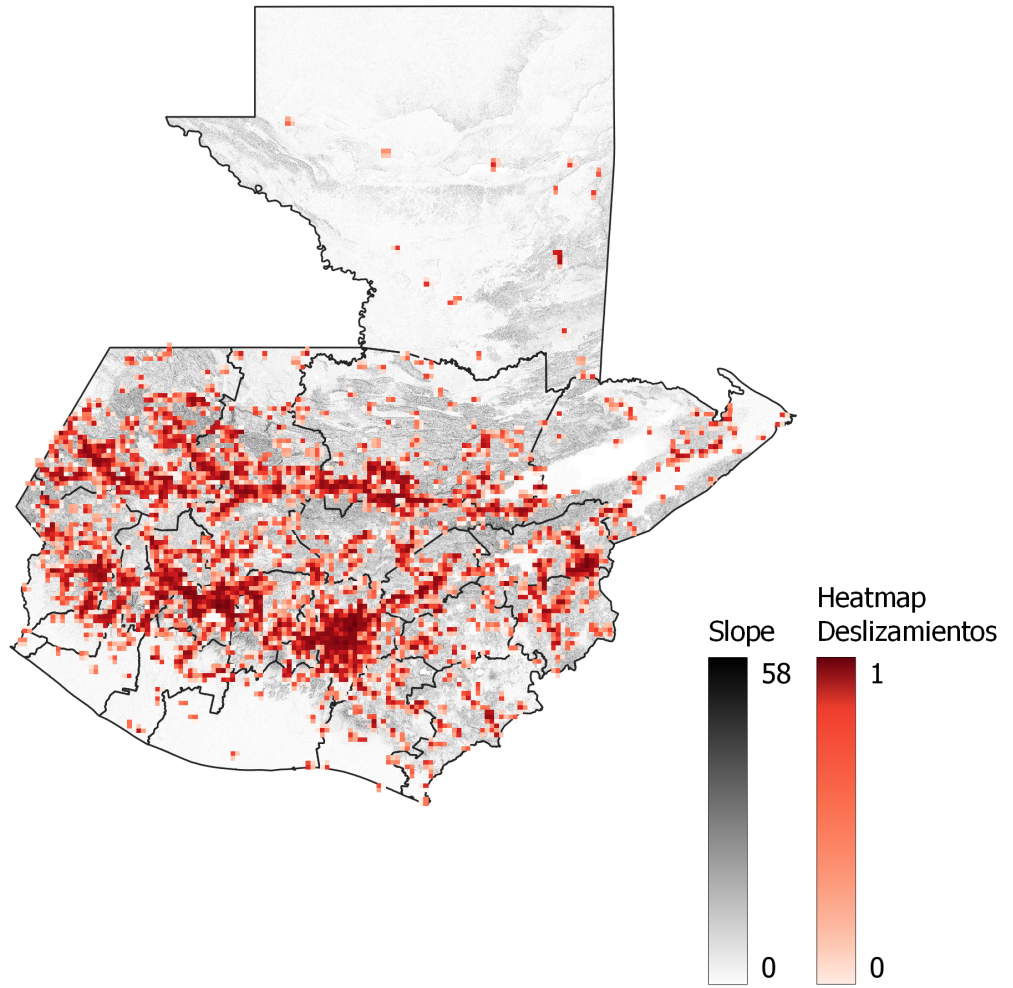
6.1. Totales de datos de inventarios de deslizamiento

Cuadro 42: Total de datos inventario deslizamientos

Tipo	Total	<i>Train</i>	<i>Test</i>
Deslizamientos	3978	2760	1218
No Deslizamientos	5764	4059	1705

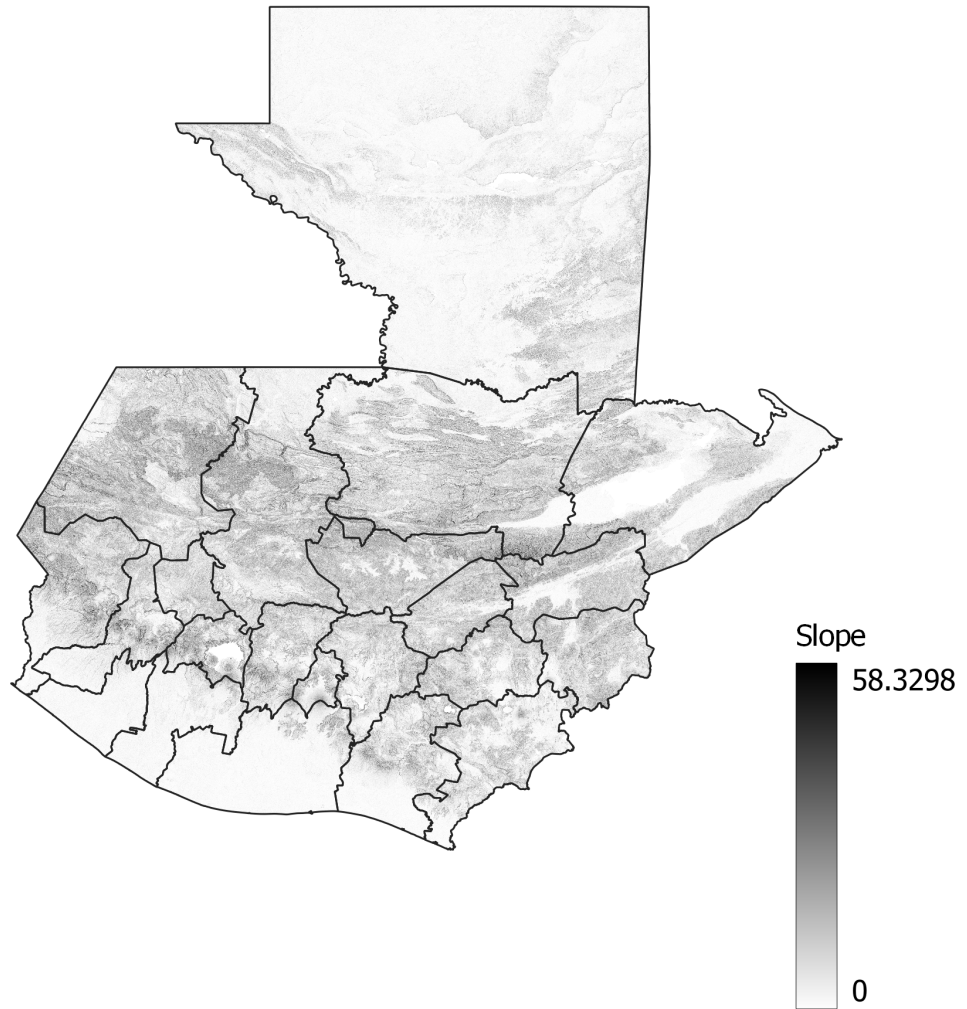
Fuente: elaboración propia

Figura 42: Mapa de Calor de deslizamientos en Guatemala



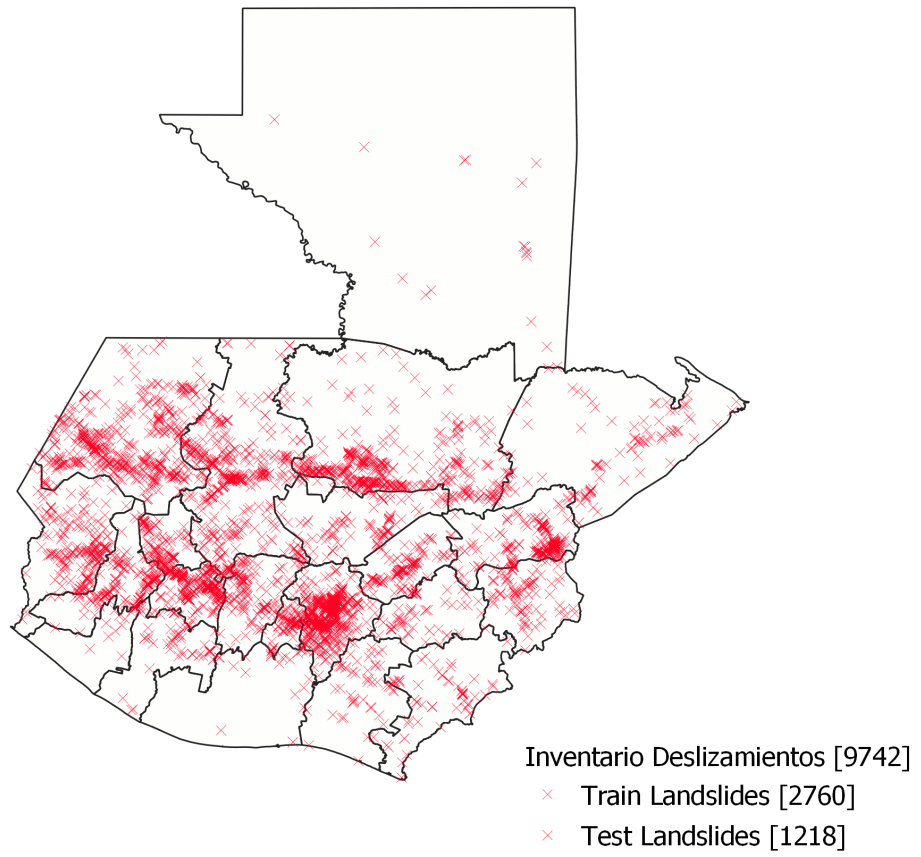
Fuente: elaboración propia

Figura 43: Mapa de Slope en Guatemala utilizando *DEM*



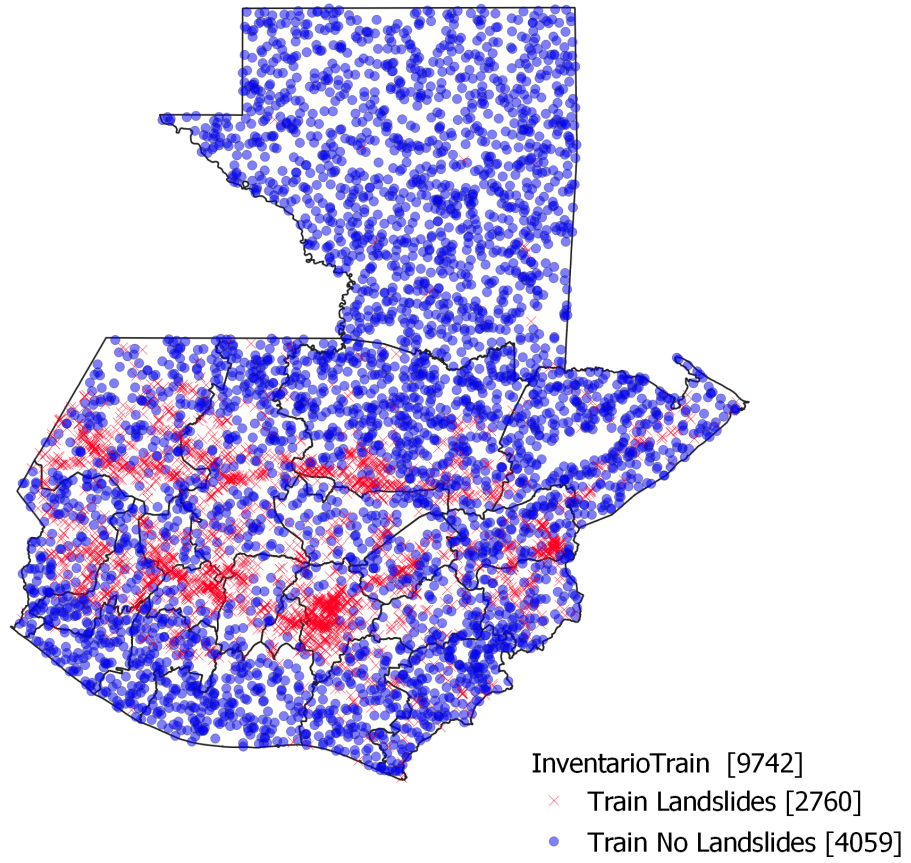
Fuente: elaboración propia

Figura 44: Mapa de deslizamientos en Guatemala



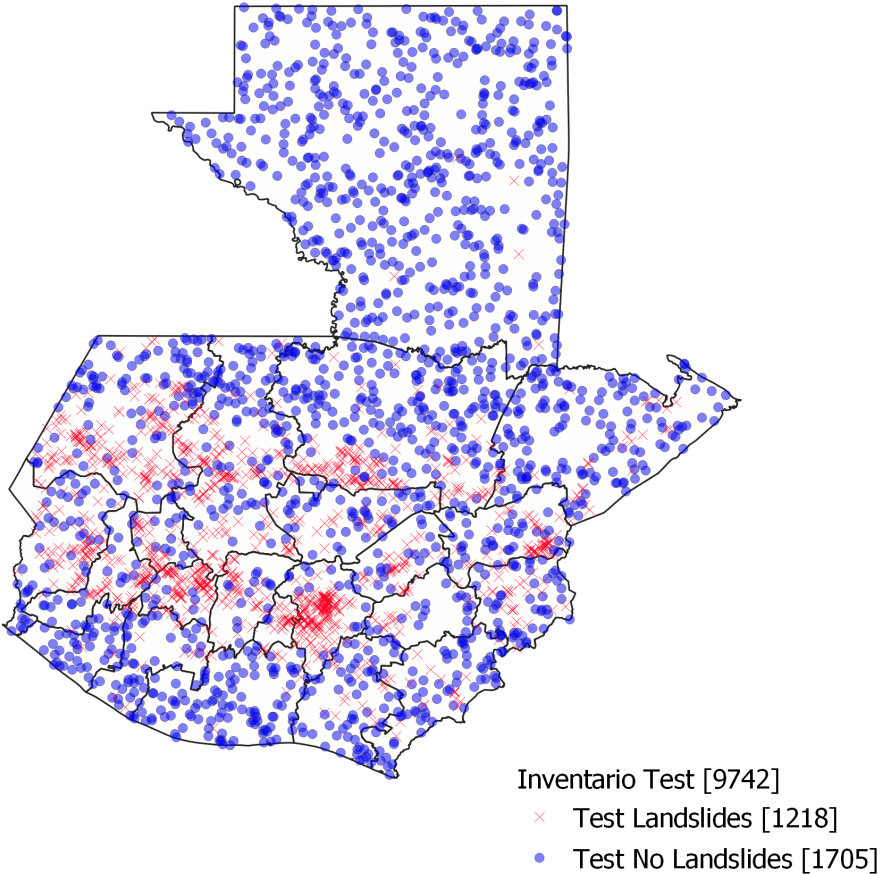
Fuente: elaboración propia

Figura 45: Mapa de *training* de deslizamientos en Guatemala



Fuente: elaboración propia

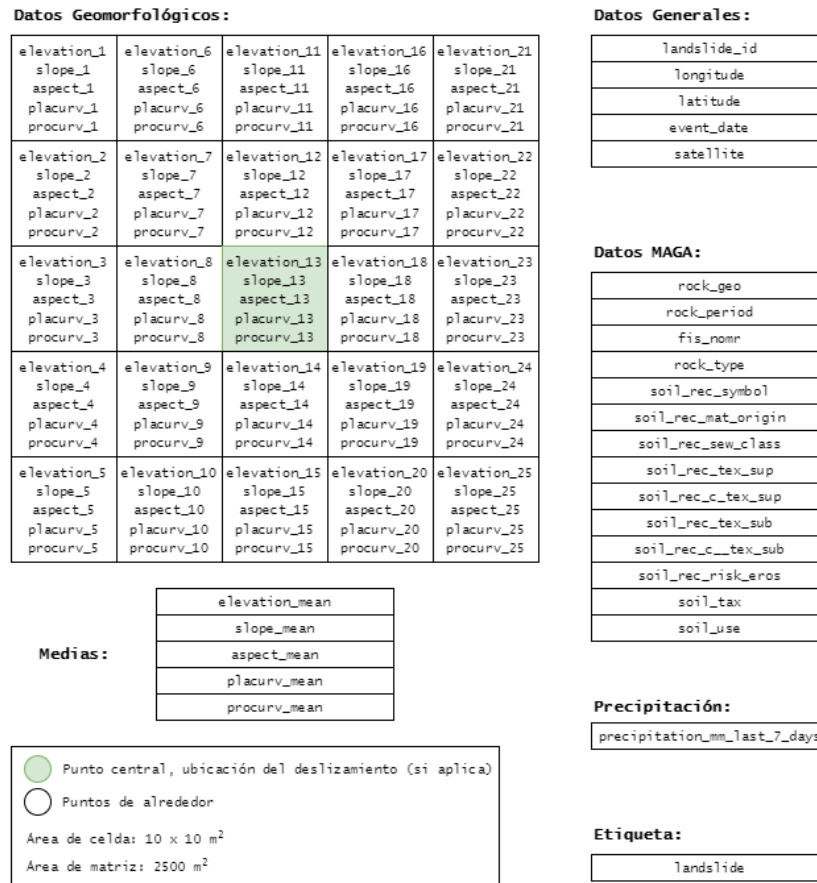
Figura 46: Mapa de testing de deslizamientos en Guatemala



Fuente: elaboración propia

6.2. Estructura del modelo

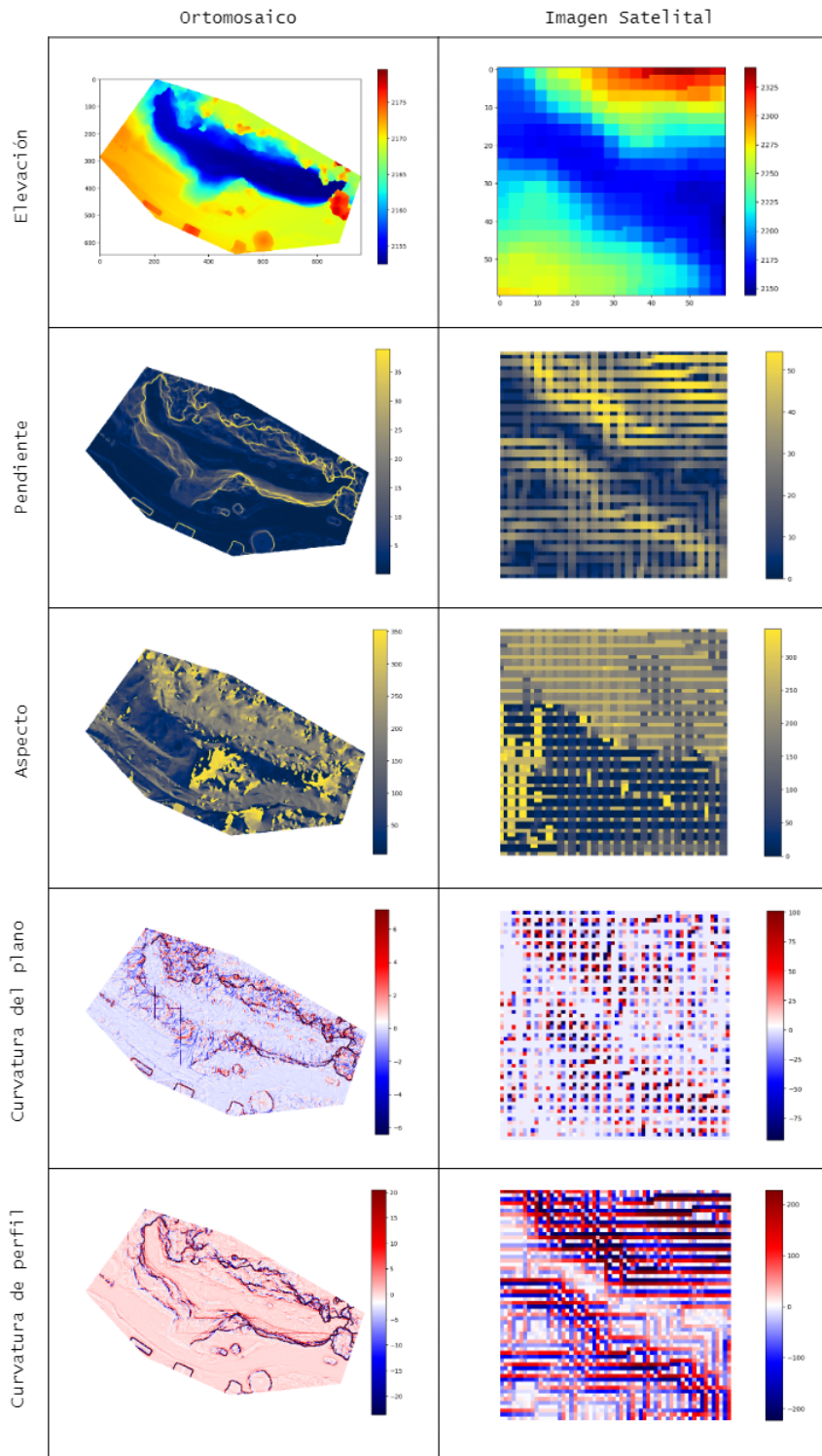
Figura 47: Diagrama del modelo de datos



Fuente: elaboración propia

6.3. Datos geomorfológicos obtenidos

Figura 48: Diagrama comparativo de datos geomorfológicos obtenidos de un ortomosaico y una imagen satelital



Fuente: elaboración propia

6.4. Resultados *ROC AUC* Y *accuracy* modelos entrenados

Cuadro 43: *Accuracy* y *roc auc* score para *train test* usando media

Algoritmo	<i>test</i>		<i>train</i>	
	<i>ROC AUC</i>	<i>Accuracy</i>	<i>ROC AUC</i>	<i>Accuracy</i>
DecistionTree	0.81	0.81	1.0	1.0
RandomForest	0.95	0.89	1.0	1.0
<i>XGBoost</i>	0.95	0.89	1.0	1.0
<i>LogisticRegression</i>	0.89	0.8	0.88	0.8
SupportVectorMachine	0.92	0.84	0.92	0.84

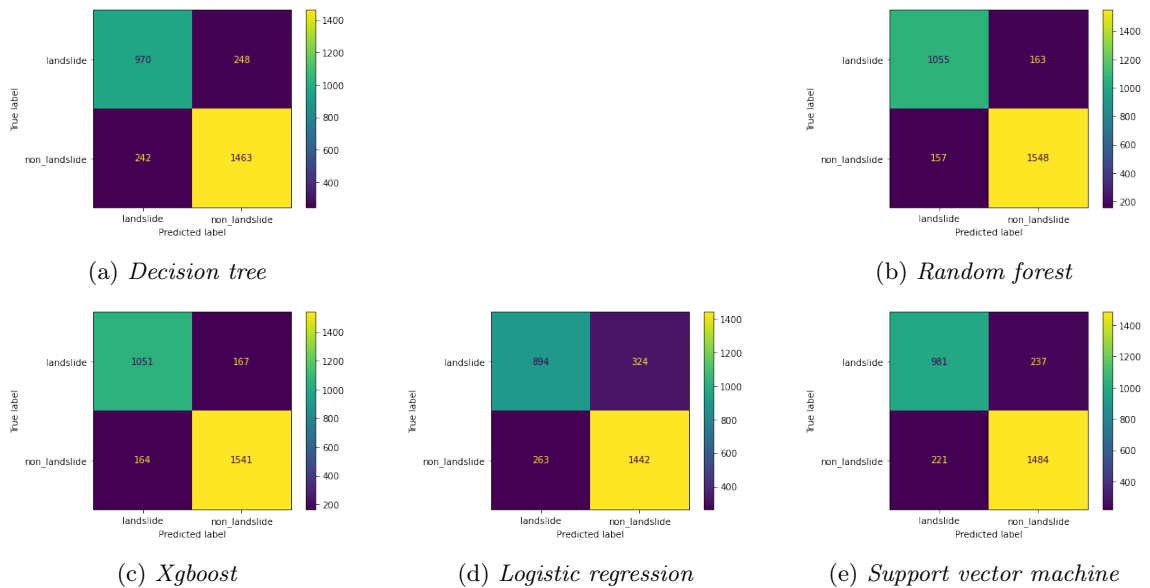
Fuente: elaboración propia

Cuadro 44: *Accuracy* y *roc auc* score para *train test* usando punto exacto

Algoritmo	<i>test</i>		<i>train</i>	
	<i>ROC AUC</i>	<i>Accuracy</i>	<i>ROC AUC</i>	<i>Accuracy</i>
DecistionTree	0.81	0.82	1.0	1.0
RandomForest	0.95	0.89	1.0	1.0
<i>XGBoost</i>	0.95	0.89	1.0	0.99
<i>LogisticRegression</i>	0.9	0.81	0.88	0.8
oo	0.92	0.85	0.92	0.84

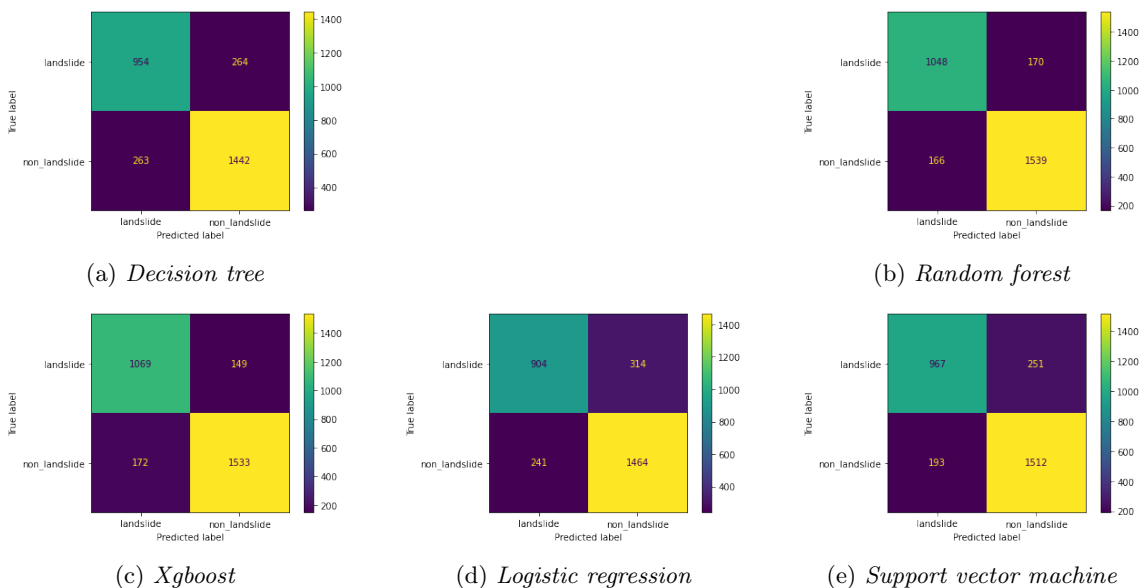
Fuente: elaboración propia

Figura 49: Matrices de confusión para modelos usando media



Fuente: elaboración propia

Figura 50: Matrices de confusión para modelos usando punto exacto

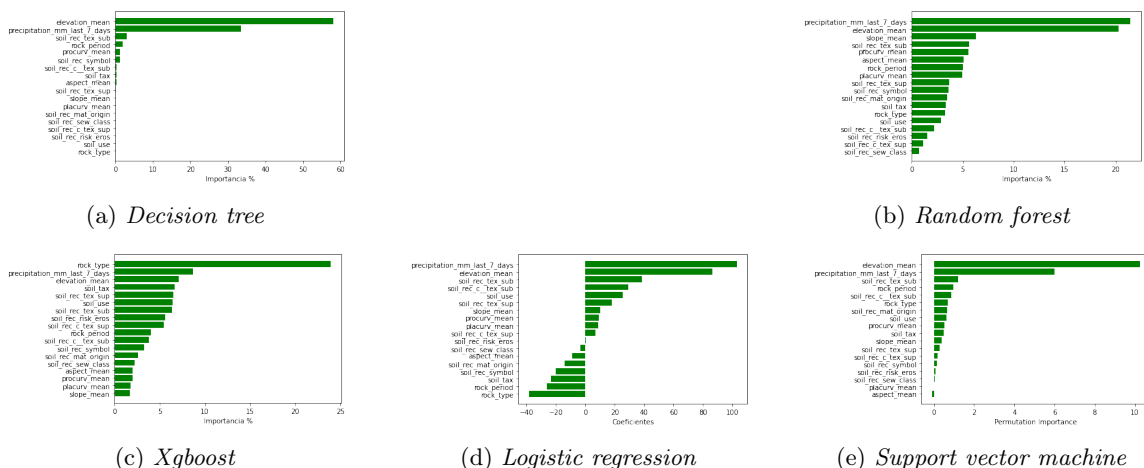


Fuente: elaboración propia

6.5. Importancia de variables de modelos

6.5.1. Importancias usando media

Figura 51: Importancia/peso de variables para modelos usando media



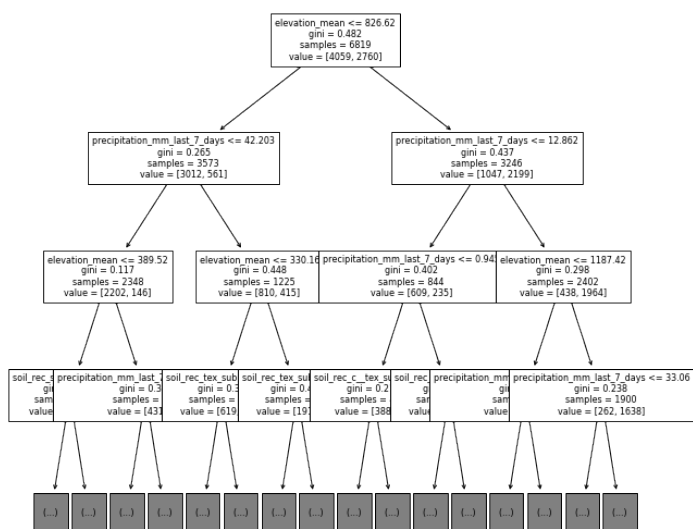
Fuente: elaboración propia

Cuadro 45: Importancia de variables para árboles de decisión usando media

No.	Variable	Importancia %
1	<i>elevation_mean</i>	58.17
2	<i>precipitation_mm_last_7_days</i>	33.49
3	<i>soil_rec_tex_sub</i>	3.11
4	<i>rock_period</i>	1.97
5	<i>procurv_mean</i>	1.25
6	<i>soil_rec_symbol</i>	1.2
7	<i>soil_rec_c_tex_sub</i>	0.24
8	<i>soil_tax</i>	0.21
9	<i>aspect_mean</i>	0.2
10	<i>soil_rec_tex_sup</i>	0.16
11	<i>slope_mean</i>	0.0
12	<i>placurv_mean</i>	0.0
13	<i>soil_rec_mat_origin</i>	0.0
14	<i>soil_rec_sew_class</i>	0.0
15	<i>soil_rec_c_tex_sup</i>	0.0
16	<i>soil_rec_risk_eros</i>	0.0
17	<i>soil_use</i>	0.0
18	<i>rock_type</i>	0.0

Fuente: elaboración propia

Figura 52: Ejemplo para árboles de decisión con *max depth 3*



Fuente: elaboración propia

Cuadro 46: Importancia de variables para *random forest* usando media

No.	Variable	Importancia %
1	<i>precipitation_mm_last_7_days</i>	21.46
2	<i>elevation_mean</i>	20.28
3	<i>slope_mean</i>	6.29
4	<i>soil_rec_tex_sub</i>	5.64
5	<i>procurv_mean</i>	5.57
6	<i>aspect_mean</i>	5.09
7	<i>rock_period</i>	5.01
8	<i>placurv_mean</i>	4.98
9	<i>soil_rec_tex_sup</i>	3.64
10	<i>soil_rec_symbol</i>	3.58
11	<i>soil_rec_mat_origin</i>	3.47
12	<i>soil_tax</i>	3.31
13	<i>rock_type</i>	3.29
14	<i>soil_use</i>	2.88
15	<i>soil_rec_c_tex_sub</i>	2.2
16	<i>soil_rec_risk_eros</i>	1.51
17	<i>soil_rec_c_tex_sup</i>	1.1

Fuente: elaboración propia

Cuadro 47: Importancia de variables para *xgboost* usando media

No.	Variable	Importancia %
1	<i>rock_type</i>	23.96
2	<i>precipitation_mm_last_7_days</i>	8.72
3	<i>elevation_mean</i>	7.07
4	<i>soil_tax</i>	6.62
5	<i>soil_rec_tex_sup</i>	6.48
6	<i>soil_use</i>	6.4
7	<i>soil_rec_tex_sub</i>	6.38
8	<i>soil_rec_risk_eros</i>	5.59
9	<i>soil_rec_c_tex_sup</i>	5.47
10	<i>rock_period</i>	4.04
11	<i>soil_rec_c_tex_sub</i>	3.8
12	<i>soil_rec_symbol</i>	3.25
13	<i>soil_rec_mat_origin</i>	2.56
14	<i>soil_rec_sew_class</i>	2.19
15	<i>aspect_mean</i>	2.03
16	<i>procurv_mean</i>	1.97
17	<i>placurv_mean</i>	1.77

Fuente: elaboración propia

Cuadro 48: Coeficientes de variables para *logisticregression* usando media

No.	Variable	Importancia %
1	<i>precipitation_mm_last_7_days</i>	103.06
2	<i>elevation_mean</i>	86.13
3	<i>soil_rec_tex_sub</i>	38.69
4	<i>soil_rec_c__tex_sub</i>	29.38
5	<i>soil_use</i>	25.32
6	<i>soil_rec_tex_sup</i>	18.18
7	<i>slope_mean</i>	10.34
8	<i>procurv_mean</i>	9.26
9	<i>placurv_mean</i>	8.83
10	<i>soil_rec_c__tex_sup</i>	6.9
11	<i>soil_rec_risk_eros</i>	0.6
12	<i>soil_rec_sew_class</i>	-3.26
13	<i>aspect_mean</i>	-8.91
14	<i>soil_rec_mat_origin</i>	-13.93
15	<i>soil_rec_symbol</i>	-20.06
16	<i>soil_tax</i>	-23.23
17	<i>rock_period</i>	-25.92

Fuente: elaboración propia

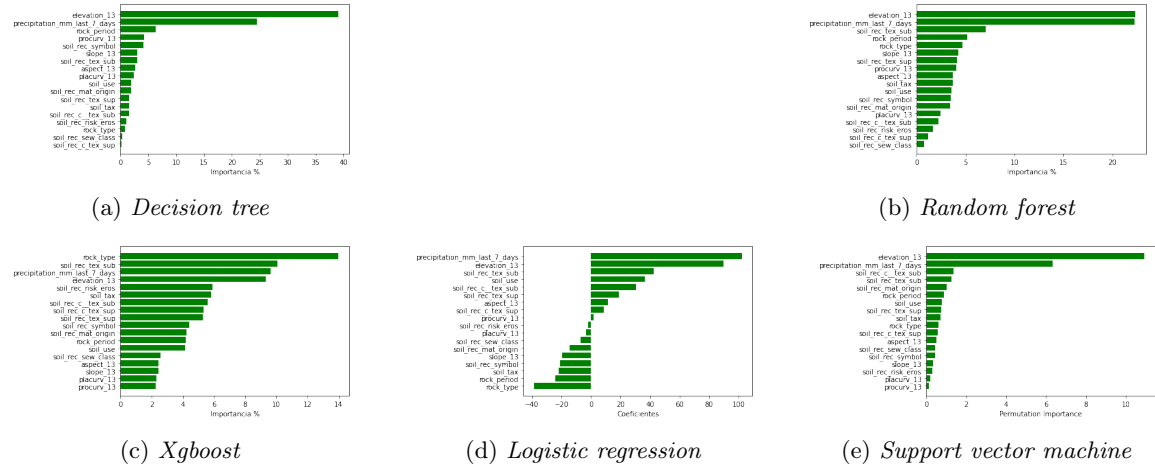
Cuadro 49: Coeficientes de variables para *supportvectormachine* usando media

No.	Variable	Importancia %
1	<i>elevation_mean</i>	10.25
2	<i>precipitation_mm_last_7_days</i>	5.99
3	<i>soil_rec_tex_sub</i>	1.19
4	<i>rock_period</i>	0.97
5	<i>soil_rec_c__tex_sub</i>	0.85
6	<i>rock_type</i>	0.7
7	<i>soil_rec_mat_origin</i>	0.66
8	<i>soil_use</i>	0.62
9	<i>procurv_mean</i>	0.51
10	<i>soil_tax</i>	0.5
11	<i>slope_mean</i>	0.39
12	<i>soil_rec_tex_sup</i>	0.29
13	<i>soil_rec_c__tex_sup</i>	0.17
14	<i>soil_rec_symbol</i>	0.14
15	<i>soil_rec_risk_eros</i>	0.09
16	<i>soil_rec_sew_class</i>	0.04
17	<i>placurv_mean</i>	0.01

Fuente: elaboración propia

6.5.2. Importancias usando punto exacto

Figura 53: Importancia/peso de variables para modelos usando punto exacto



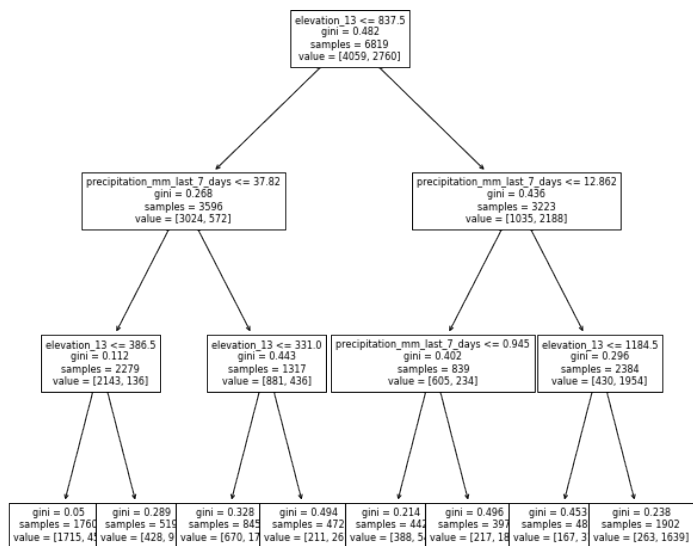
Fuente: elaboración propia

Cuadro 50: Importancia de variables para árboles de decisión usando punto exacto

No.	Variable	Importancia %
1	<i>elevation_13</i>	39.14
2	<i>precipitation_mm_last_7_days</i>	24.53
3	<i>rock_period</i>	6.27
4	<i>procurv_13</i>	4.2
5	<i>soil_rec_symbol</i>	4.15
6	<i>slope_13</i>	3.0
7	<i>soil_rec_tex_sub</i>	3.0
8	<i>aspect_13</i>	2.68
9	<i>placurv_13</i>	2.37
10	<i>soil_use</i>	1.9
11	<i>soil_rec_mat_origin</i>	1.88
12	<i>soil_rec_tex_sup</i>	1.6
13	<i>soil_tax</i>	1.52
14	<i>soil_rec_c__tex_sub</i>	1.48
15	<i>soil_rec_risk_eros</i>	1.0
16	<i>rock_type</i>	0.75
17	<i>soil_rec_sew_class</i>	0.33
18	<i>soil_rec_c_tex_sup</i>	0.21

Fuente: elaboración propia

Figura 54: Ejemplo para árboles de decisión con *max depth* 3



Fuente: elaboración propia

Cuadro 51: Importancia de variables para *random forest* usando punto exacto

No.	Variable	Importancia %
1	<i>elevation_13</i>	22.37
2	<i>precipitation_mm_last_7_days</i>	22.26
3	<i>soil_rec_tex_sub</i>	7.07
4	<i>rock_period</i>	5.18
5	<i>rock_type</i>	4.69
6	<i>slope_13</i>	4.27
7	<i>soil_rec_tex_sup</i>	4.14
8	<i>procurv_13</i>	4.02
9	<i>aspect_13</i>	3.67
10	<i>soil_tax</i>	3.66
11	<i>soil_use</i>	3.58
12	<i>soil_rec_symbol</i>	3.51
13	<i>soil_rec_mat_origin</i>	3.42
14	<i>placurv_13</i>	2.42
15	<i>soil_rec_c__tex_sub</i>	2.24
16	<i>soil_rec_risk_eros</i>	1.67
17	<i>soil_rec_c__tex_sup</i>	1.13

Fuente: elaboración propia

Cuadro 52: Importancia de variables para *xgboost* usando punto exacto

No.	Variable	Importancia %
1	<i>rock_type</i>	13.99
2	<i>soil_rec_tex_sub</i>	10.05
3	<i>precipitation_mm_last_7_days</i>	9.65
4	<i>elevation_13</i>	9.34
5	<i>soil_rec_risk_eros</i>	5.91
6	<i>soil_tax</i>	5.83
7	<i>soil_rec_c_tex_sub</i>	5.58
8	<i>soil_rec_c_tex_sup</i>	5.34
9	<i>soil_rec_tex_sup</i>	5.31
10	<i>soil_rec_symbol</i>	4.42
11	<i>soil_rec_mat_origin</i>	4.23
12	<i>rock_period</i>	4.19
13	<i>soil_use</i>	4.16
14	<i>soil_rec_sew_class</i>	2.55
15	<i>aspect_13</i>	2.45
16	<i>slope_13</i>	2.43
17	<i>placurv_13</i>	2.3

Fuente: elaboración propia

Cuadro 53: Coeficientes de variables para *logisticregression* usando punto exacto

No.	Variable	Importancia %
1	<i>precipitation_mm_last_7_days</i>	102.24
2	<i>elevation_13</i>	89.73
3	<i>soil_rec_tex_sub</i>	42.58
4	<i>soil_use</i>	36.5
5	<i>soil_rec_c_tex_sub</i>	30.37
6	<i>soil_rec_tex_sup</i>	18.86
7	<i>aspect_13</i>	11.47
8	<i>soil_rec_c_tex_sup</i>	8.67
9	<i>procurv_13</i>	1.73
10	<i>soil_rec_risk_eros</i>	-1.96
11	<i>placurv_13</i>	-3.38
12	<i>soil_rec_sew_class</i>	-7.25
13	<i>soil_rec_mat_origin</i>	-14.65
14	<i>slope_13</i>	-19.56
15	<i>soil_rec_symbol</i>	-21.06
16	<i>soil_tax</i>	-21.81
17	<i>rock_period</i>	-24.22

Fuente: elaboración propia

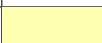
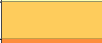



Cuadro 54: Coeficientes de variables para supportvectormachine usando punto exacto

No.	Variable	Importancia %
1	<i>elevation_13</i>	10.94
2	<i>precipitation_mm_last_7_days</i>	6.32
3	<i>soil_rec_c_tex_sub</i>	1.35
4	<i>soil_rec_tex_sub</i>	1.24
5	<i>soil_rec_mat_origin</i>	1.03
6	<i>rock_period</i>	0.88
7	<i>soil_use</i>	0.79
8	<i>soil_rec_tex_sup</i>	0.73
9	<i>soil_tax</i>	0.7
10	<i>rock_type</i>	0.61
11	<i>soil_rec_c_tex_sup</i>	0.57
12	<i>aspect_13</i>	0.5
13	<i>soil_rec_sew_class</i>	0.44
14	<i>soil_rec_symbol</i>	0.42
15	<i>slope_13</i>	0.34
16	<i>soil_rec_risk_eros</i>	0.29
17	<i>placurv_13</i>	0.18

Fuente: elaboración propia

6.6. Validación de modelos con *Landslide Density* y *R index*

Cuadro 55: Clases de susceptibilidad para evaluación utilizando intervalos iguales

Clase	Rango %	Color
Muy Baja	0-20	
Baja	21-40	
Media	41-60	
Alta	61-80	
Muy Alta	81-100	

Fuente: elaboración propia

6.6.1. *Landslide Density* usando media

Cuadro 56: *Landslide density* usando media *validation decision tree*

Clase	% Pixeles	% Landslides <i>train</i>	% Landslides <i>test</i>	<i>LD train</i>	<i>LD test</i>
Muy Baja	50.84	7.32	7.8	0.14	0.15
Baja	12.71	6.74	7.39	0.53	0.58
Media	2.42	8.52	8.78	3.52	3.63
Alta	10.68	19.71	18.97	1.85	1.78
Muy Alta	23.35	57.72	57.06	2.47	2.44

Fuente: elaboración propia

Cuadro 57: *Landslide density* usando media *validation random forest*

Clase	% Pixeles	% Landslides <i>train</i>	% Landslides <i>test</i>	<i>LD train</i>	<i>LD test</i>
Muy Baja	51.5	1.36	4.43	0.03	0.09
Baja	14.78	1.56	5.09	0.11	0.34
Media	12.38	3.09	10.1	0.25	0.82
Alta	11.17	5.51	17.98	0.49	1.61
Muy Alta	10.16	88.49	62.4	8.71	6.14

Fuente: elaboración propia

Cuadro 58: *Landslide density* usando media *validation ada boost*

Clase	% Pixeles	% Landslides <i>train</i>	% Landslides <i>test</i>	<i>LD train</i>	<i>LD test</i>
Muy Baja	0	0	0	0	0
Baja	0	0	0	0	0
Media	100.0	100.0	100.0	1.0	1.0
Alta	0	0	0	0	0
Muy Alta	0	0	0	0	0

Fuente: elaboración propia

Cuadro 59: *Landslide density* usando media *validation xgboost*

Clase	% Pixeles	% Landslides <i>train</i>	% Landslides <i>test</i>	<i>LD train</i>	<i>LD test</i>
Muy Baja	61.39	2.34	7.64	0.04	0.12
Baja	8.38	1.21	3.94	0.14	0.47
Media	6.56	1.68	5.5	0.26	0.84
Alta	7.11	3.8	9.85	0.53	1.39
Muy Alta	16.56	90.98	73.07	5.49	4.41

Fuente: elaboración propia

Cuadro 60: *Landslide density* usando media *validation logistic regression*

Clase	% Pixeles	% Landslides <i>train</i>	% Landslides <i>test</i>	<i>LD train</i>	<i>LD test</i>
Muy Baja	48.26	4.85	4.27	0.1	0.09
Baja	17.21	12.37	12.64	0.72	0.73
Media	14.08	19.48	19.87	1.38	1.41
Alta	14.65	27.02	27.75	1.84	1.89
Muy Alta	5.8	36.27	35.47	6.25	6.11

Fuente: elaboración propia

Cuadro 61: *Landslide density* usando media *validation support vector machine*

Clase	% Pixeles	% Landslides <i>train</i>	% Landslides <i>test</i>	<i>LD train</i>	<i>LD test</i>
Muy Baja	55.04	7.26	6.9	0.13	0.13
Baja	15.09	9.05	8.87	0.6	0.59
Media	10.2	10.43	10.1	1.02	0.99
Alta	10.07	19.71	21.51	1.96	2.14
Muy Alta	9.59	53.54	52.63	5.58	5.48

Fuente: elaboración propia

6.6.2. *Landslide Density* usando punto exacto

Cuadro 62: *Landslide density* usando punto exacto *validation decision tree*

Clase	% Pixeles	% Landslides <i>train</i>	% Landslides <i>test</i>	<i>LD train</i>	<i>LD test</i>
Muy Baja	14.13	6.84	6.65	0.48	0.47
Baja	38.68	6.1	5.5	0.16	0.14
Media	14.59	15.65	15.02	1.07	1.03
Alta	7.85	11.3	11.08	1.44	1.41
Muy Alta	24.75	60.11	61.74	2.43	2.49

Fuente: elaboración propia

Cuadro 63: *Landslide density* usando punto exacto *validation random forest*

Clase	% Pixeles	% Landslides <i>train</i>	% Landslides <i>test</i>	<i>LD train</i>	<i>LD test</i>
Muy Baja	50.68	1.34	4.35	0.03	0.09
Baja	14.62	1.6	5.17	0.11	0.35
Media	13.33	3.62	10.92	0.27	0.82
Alta	12.07	14.41	18.31	1.19	1.52
Muy Alta	9.3	79.03	61.25	8.5	6.59

Fuente: elaboración propia

Cuadro 64: *Landslide density* usando punto exacto *validation ada boost*

Clase	% Pixeles	% Landslides <i>train</i>	% Landslides <i>test</i>	<i>LD train</i>	<i>LD test</i>
Muy Baja	0	0	0	0	0
Baja	0	0	0	0	0
Media	100.0	100.0	100.0	1.0	1.0
Alta	0	0	0	0	0
Muy Alta	0	0	0	0	0

Fuente: elaboración propia

Cuadro 65: *Landslide density* usando punto exacto *validation xgboost*

Clase	% Pixeles	% Landslides <i>train</i>	% Landslides <i>test</i>	<i>LD train</i>	<i>LD test</i>
Muy Baja	58.1	1.8	5.83	0.03	0.1
Baja	10.78	1.52	4.35	0.14	0.4
Media	8.7	3.27	6.9	0.38	0.79
Alta	9.62	10.03	13.38	1.04	1.39
Muy Alta	12.79	83.38	69.54	6.52	5.44

Fuente: elaboración propia

Cuadro 66: *Landslide density* usando punto exacto *validation logistic regression*

Clase	% Pixeles	% Landslides <i>train</i>	% Landslides <i>test</i>	<i>LD train</i>	<i>LD test</i>
Muy Baja	48.85	5.29	4.27	0.11	0.09
Baja	18.31	12.18	13.38	0.67	0.73
Media	14.16	18.24	17.82	1.29	1.26
Alta	13.36	26.11	25.62	1.95	1.92
Muy Alta	5.31	38.17	38.92	7.19	7.33

Fuente: elaboración propia

Cuadro 67: *Landslide density* usando punto exacto *validation support vector machine*

Clase	% Pixeles	% Landslides <i>train</i>	% Landslides <i>test</i>	<i>LD train</i>	<i>LD test</i>
Muy Baja	55.01	6.69	6.16	0.12	0.11
Baja	15.96	9.95	8.87	0.62	0.56
Media	10.6	11.17	11.74	1.05	1.11
Alta	9.23	17.65	18.47	1.91	2.0
Muy Alta	9.2	54.53	54.76	5.93	5.95

Fuente: elaboración propia

6.6.3. *Landslide R Index* usando media

Cuadro 68: *Landslide r index* usando media validation decision tree

Clase	ni	Ni	ni/Ni	<i>LD R-index</i>
Muy Baja	291	49302	0.005902	0.00016907
Baja	268	12325	0.021744	0.00062285
Media	339	2348	0.144378	0.0041356
Alta	784	10355	0.075712	0.00216872
Muy Alta	2296	22649	0.101373	0.00290376

Fuente: elaboración propia

Cuadro 69: *Landslide r index* usando media validation random forest

Clase	ni	Ni	ni/Ni	<i>LD R-index</i>
Muy Baja	54	49947	0.001081	2.751e-05
Baja	62	14337	0.004324	0.00011004
Media	123	12010	0.010241	0.00026061
Alta	219	10828	0.020225	0.00051467
Muy Alta	3520	9857	0.357107	0.00908717

Fuente: elaboración propia

Cuadro 70: *Landslide r index* usando media validation ada boost

Clase	ni	Ni	ni/Ni	<i>LD R-index</i>
Muy Baja	0	0	0	0.0
Baja	0	0	0	0.0
Media	3978	96979	0.041019	0.01
Alta	0	0	0	0.0
Muy Alta	0	0	0	0.0

Fuente: elaboración propia

Cuadro 71: *Landslide r index* usando media validation xgboost

Clase	ni	Ni	ni/Ni	<i>LD R-index</i>
Muy Baja	93	59537	0.001562	5.89e-05
Baja	48	8127	0.005906	0.00022271
Media	67	6358	0.010538	0.00039736
Alta	151	6893	0.021906	0.00082603
Muy Alta	3619	16064	0.225286	0.008495

Fuente: elaboración propia

Cuadro 72: *Landslide r index* usando *media validation logistic regression*

Clase	ni	Ni	ni/Ni	<i>LD R-index</i>
Muy Baja	193	46803	0.004124	9.76e-05
Baja	492	16691	0.029477	0.00069764
Media	775	13654	0.05676	0.00134335
Alta	1075	14205	0.075678	0.00179108
Muy Alta	1443	5626	0.256488	0.00607034

Fuente: elaboración propia

Cuadro 73: *Landslide r index* usando *media validation support vector machine*

Clase	ni	Ni	ni/Ni	<i>LD R-index</i>
Muy Baja	289	53382	0.005414	0.00014205
Baja	360	14630	0.024607	0.00064563
Media	415	9892	0.041953	0.00110076
Alta	784	9770	0.080246	0.00210547
Muy Alta	2130	9305	0.228909	0.00600609

Fuente: elaboración propia

6.6.4. *Landslide R Index* usando punto exacto

Cuadro 74: *Landslide r index* usando punto exacto *validation decision tree*

Clase	ni	Ni	ni/Ni	<i>LD R-index</i>
Muy Baja	270	13703	0.019704	0.0008671
Baja	241	37508	0.006425	0.00028276
Media	618	14151	0.043672	0.00192187
Alta	446	7615	0.058569	0.00257743
Muy Alta	2373	24002	0.098867	0.00435084

Fuente: elaboración propia

Cuadro 75: *Landslide r index* usando punto exacto *validation random forest*

Clase	ni	Ni	ni/Ni	<i>LD R-index</i>
Muy Baja	53	49153	0.001078	2.622e-05
Baja	63	14176	0.004444	0.00010808
Media	143	12927	0.011062	0.00026904
Alta	569	11705	0.048612	0.00118227
Muy Alta	3120	9018	0.345975	0.00841438

Fuente: elaboración propia

Cuadro 76: *Landslide r index* usando punto exacto *validation ada boost*

Clase	ni	Ni	ni/Ni	<i>LD R-index</i>
Muy Baja	0	0	0	0.0
Baja	0	0	0	0.0
Media	3948	96979	0.04071	0.01
Alta	0	0	0	0.0
Muy Alta	0	0	0	0.0

Fuente: elaboración propia

Cuadro 77: *Landslide r index* usando punto exacto *validation xgboost*

Clase	ni	Ni	ni/Ni	<i>LD R-index</i>
Muy Baja	71	56346	0.00126	3.817e-05
Baja	60	10459	0.005737	0.00017379
Media	129	8437	0.01529	0.00046319
Alta	396	9332	0.042435	0.00128552
Muy Alta	3292	12405	0.265377	0.00803933

Fuente: elaboración propia

Cuadro 78: *Landslide r index* usando punto exacto *validation logistic regression*

Clase	ni	Ni	ni/Ni	<i>LD R-index</i>
Muy Baja	209	47373	0.004412	9.673e-05
Baja	481	17759	0.027085	0.00059386
Media	720	13736	0.052417	0.00114929
Alta	1031	12961	0.079546	0.00174413
Muy Alta	1507	5150	0.292621	0.00641599

Fuente: elaboración propia

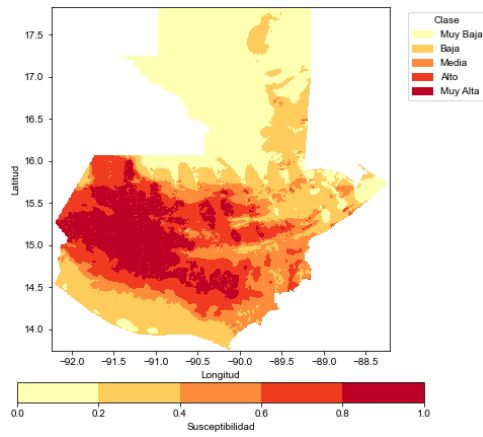
Cuadro 79: *Landslide r index* usando punto exacto *validation support vector machine*

Clase	ni	Ni	ni/Ni	<i>LD R-index</i>
Muy Baja	264	53351	0.004948	0.00012609
Baja	393	15481	0.025386	0.00064688
Media	441	10277	0.042911	0.00109346
Alta	697	8947	0.077903	0.00198512
Muy Alta	2153	8923	0.241287	0.00614844

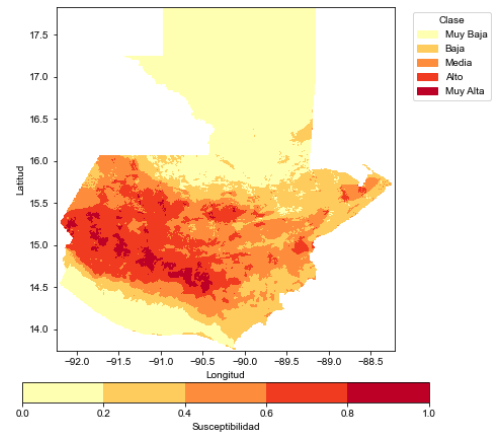
Fuente: elaboración propia

6.7. Mapas de susceptibilidad generados por los modelos

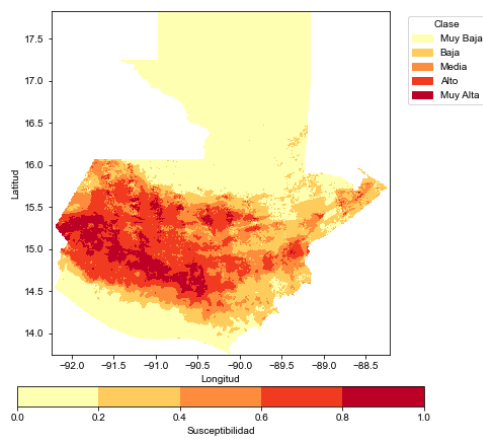
Figura 55: Mapa de susceptibilidad para modelos usando media



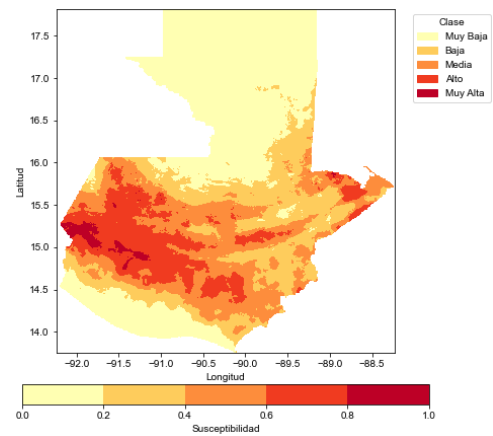
(a) Mapa de susceptibilidad *decision tree*



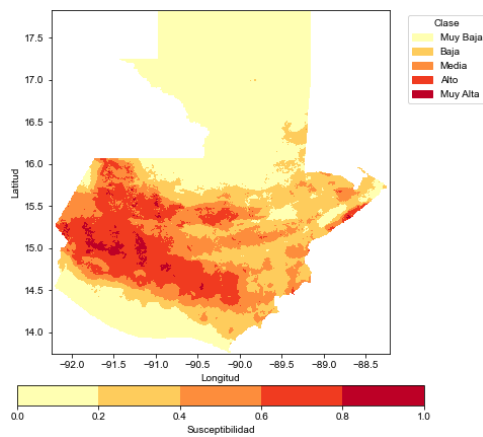
(b) Mapa de susceptibilidad *random forest*



(c) Mapa de susceptibilidad *xgboost*



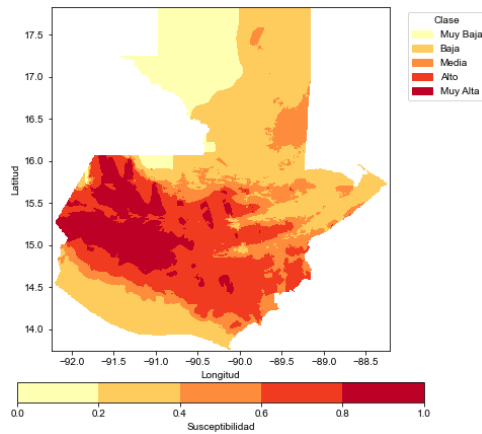
(d) Mapa de susceptibilidad *logistic regression*



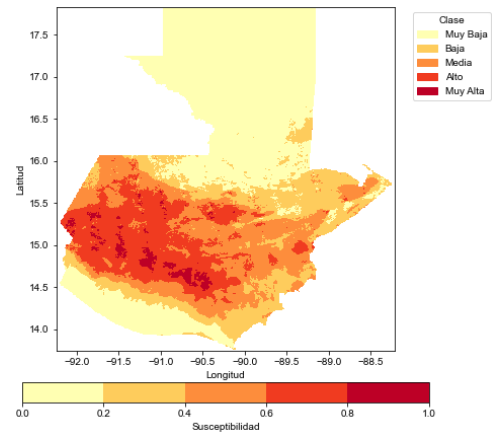
(e) Mapa de susceptibilidad *support vector machine*

Fuente: elaboración propia

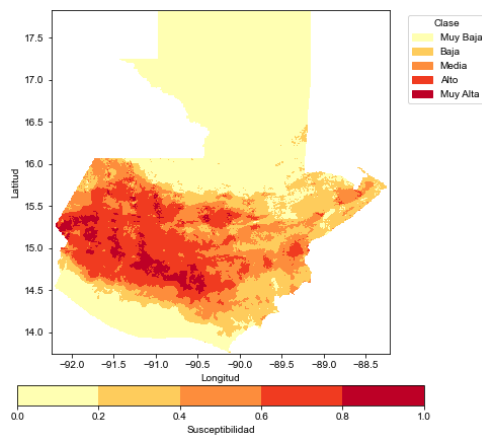
Figura 56: Mapa de susceptibilidad para modelos usando punto exacto



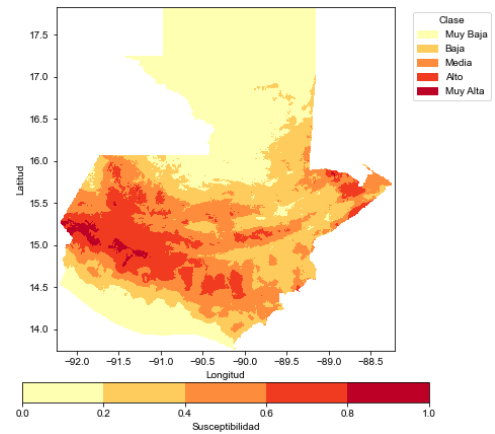
(a) Mapa de susceptibilidad *decision tree*



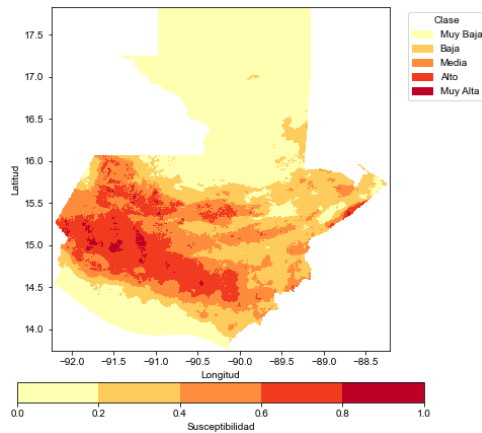
(b) Mapa de susceptibilidad *random forest*



(c) Mapa de susceptibilidad *xgboost*



(d) Mapa de susceptibilidad *logistic regression*



(e) Mapa de susceptibilidad *support vector machine*

Fuente: elaboración propia

6.8. Resumen de resultados y aprobación de modelos

Cuadro 80: Resultados generales usando media

Algoritmo	Rel Error	<i>ROC</i>	<i>Accuracy</i>	Sensitivity	<i>Specifity</i>	F1Score	<i>LD</i>	<i>R-index</i>	Appr
DecistionTree	0.429	0.81	0.81	0.84	0.78	0.78	X	X	X
RandomForest	0.376	0.95	0.89	0.91	0.87	0.87	✓	✓	✓
<i>XGBoost</i>	0.269	0.95	0.89	0.9	0.86	0.86	✓	✓	✓
<i>LogisticRegression</i>	0.5	0.89	0.8	0.85	0.73	0.75	✓	✓	✓
SupportVectorMachine	0.474	0.92	0.84	0.87	0.81	0.81	✓	✓	✓

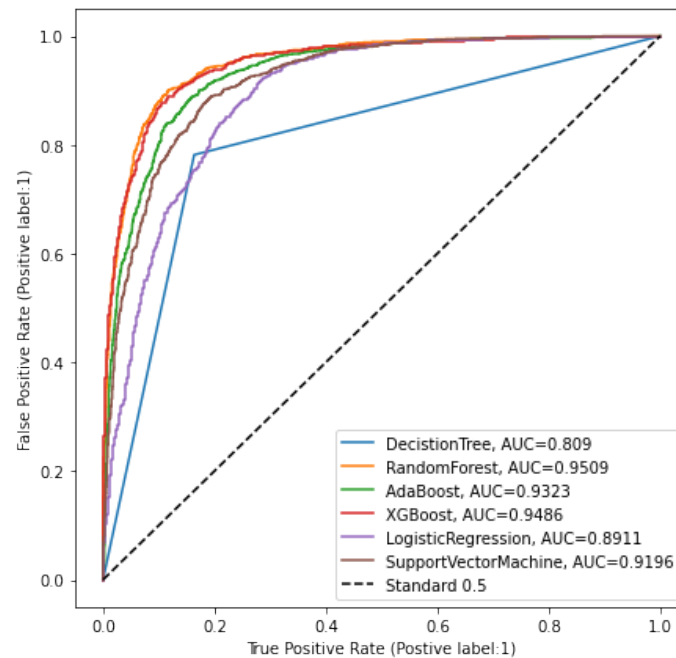
Fuente: elaboración propia

Cuadro 81: Resultados generales usando punto exacto

Algoritmo	Rel Error	<i>ROC</i>	<i>Accuracy</i>	Sensitivity	<i>Specifity</i>	F1Score	<i>LD</i>	<i>R-index</i>	Appr
DecistionTree	0.383	0.81	0.82	0.85	0.78	0.78	✓	X	X
RandomForest	0.388	0.95	0.89	0.9	0.86	0.86	✓	✓	✓
<i>XGBoost</i>	0.305	0.95	0.89	0.9	0.87	0.87	✓	✓	✓
<i>LogisticRegression</i>	0.5	0.9	0.81	0.86	0.74	0.77	✓	✓	✓
SupportVectorMachine	0.452	0.92	0.85	0.89	0.79	0.81	✓	✓	✓

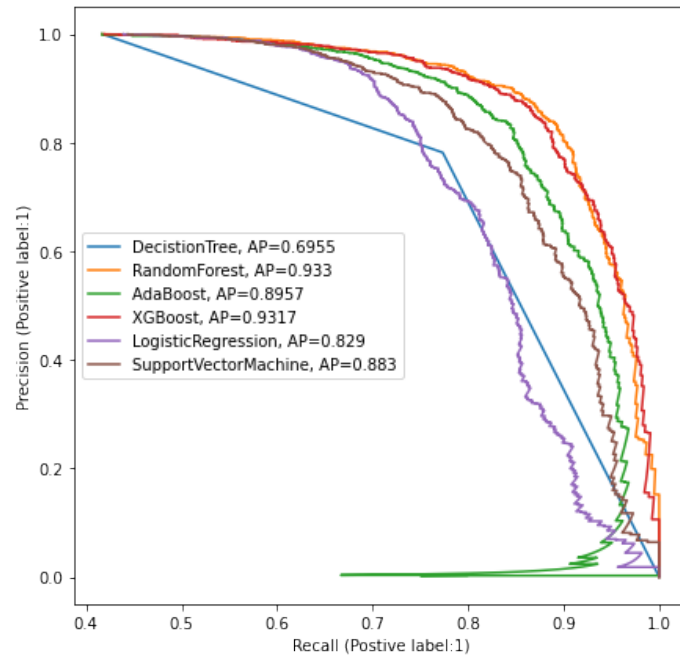
Fuente: elaboración propia

Figura 57: Roc auc curve scores usando media



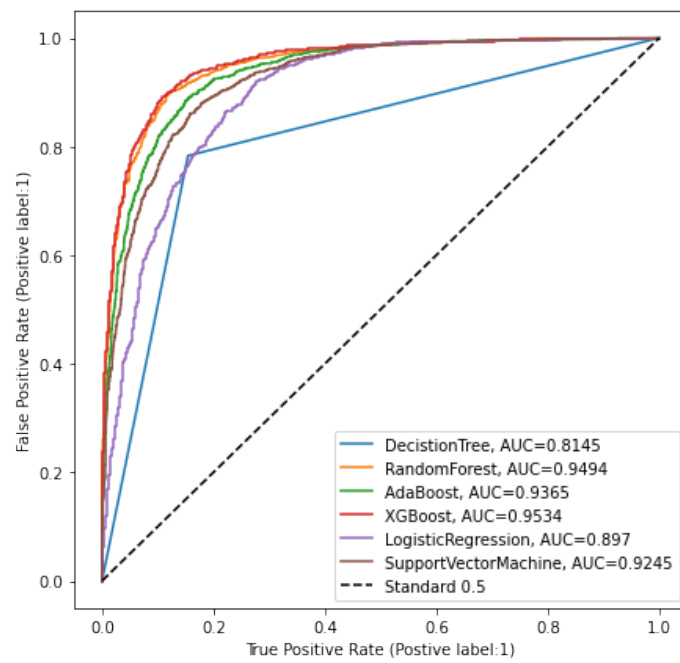
Fuente: elaboración propia

Figura 58: *Precision vs recall curve scores* usando media



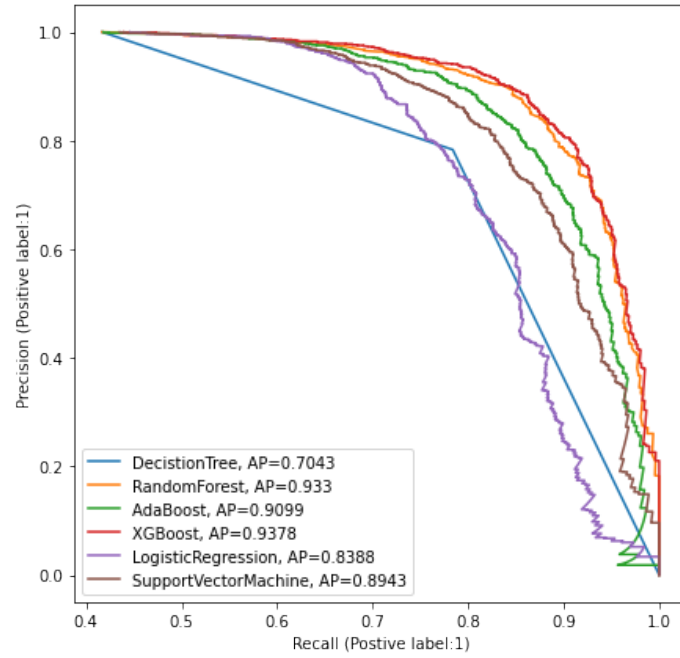
Fuente: elaboración propia

Figura 59: *Roc auc curve scores* usando punto exacto



Fuente: elaboración propia

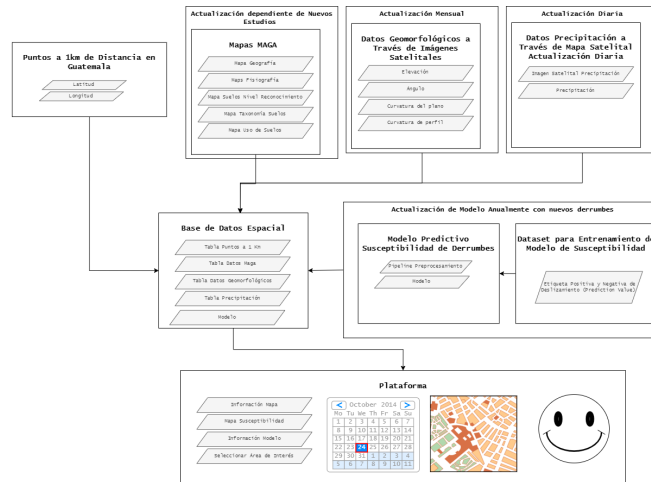
Figura 60: Precision vs recall curve scores usando punto exacto



Fuente: elaboración propia

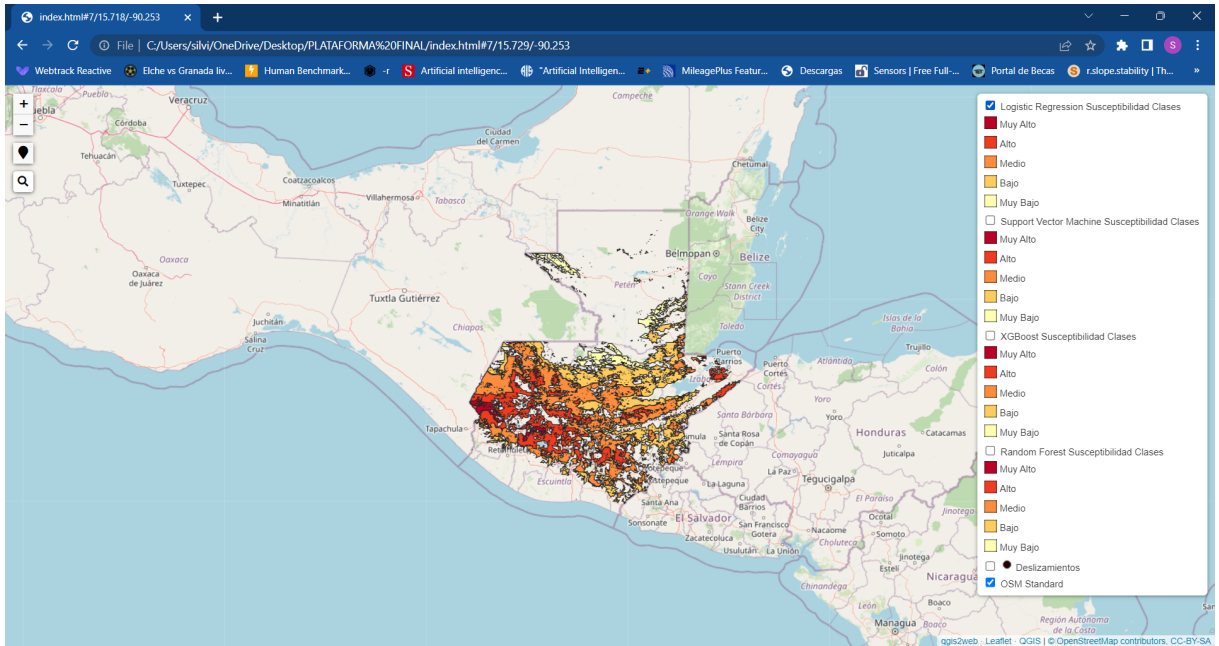
6.9. Plataforma de susceptibilidad de deslizamientos

Figura 61: Diagrama de diseño del sistema de monitorización de derrumbes



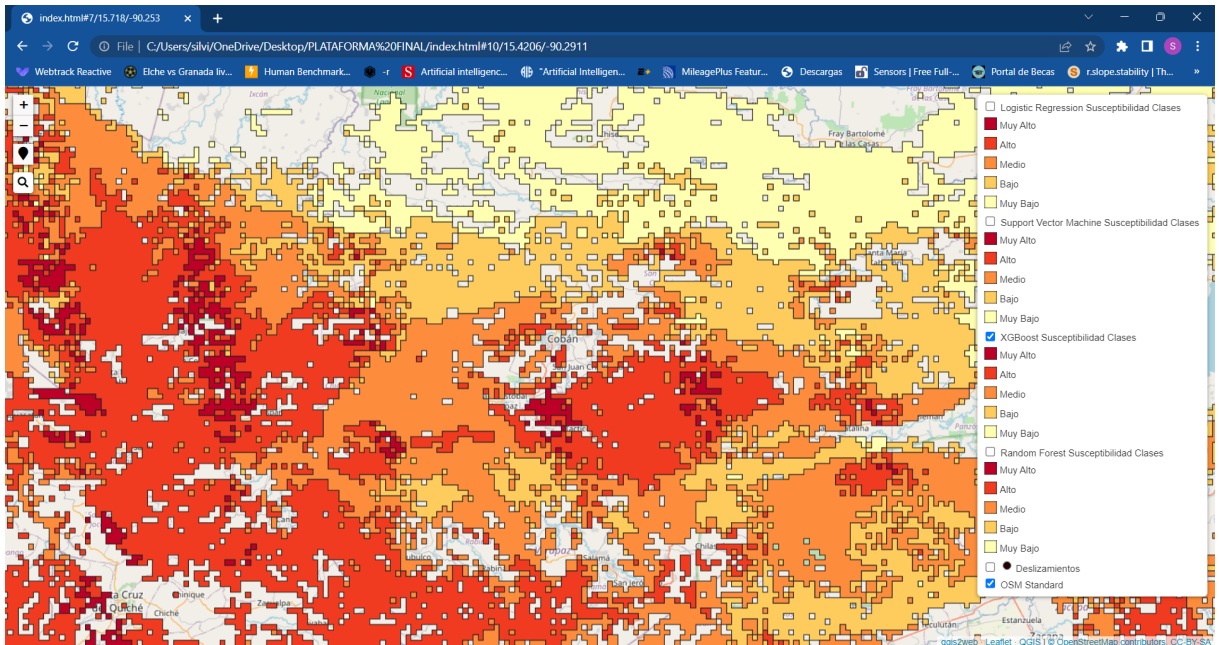
Fuente: elaboración propia

Figura 62: Imagen de plataforma 1



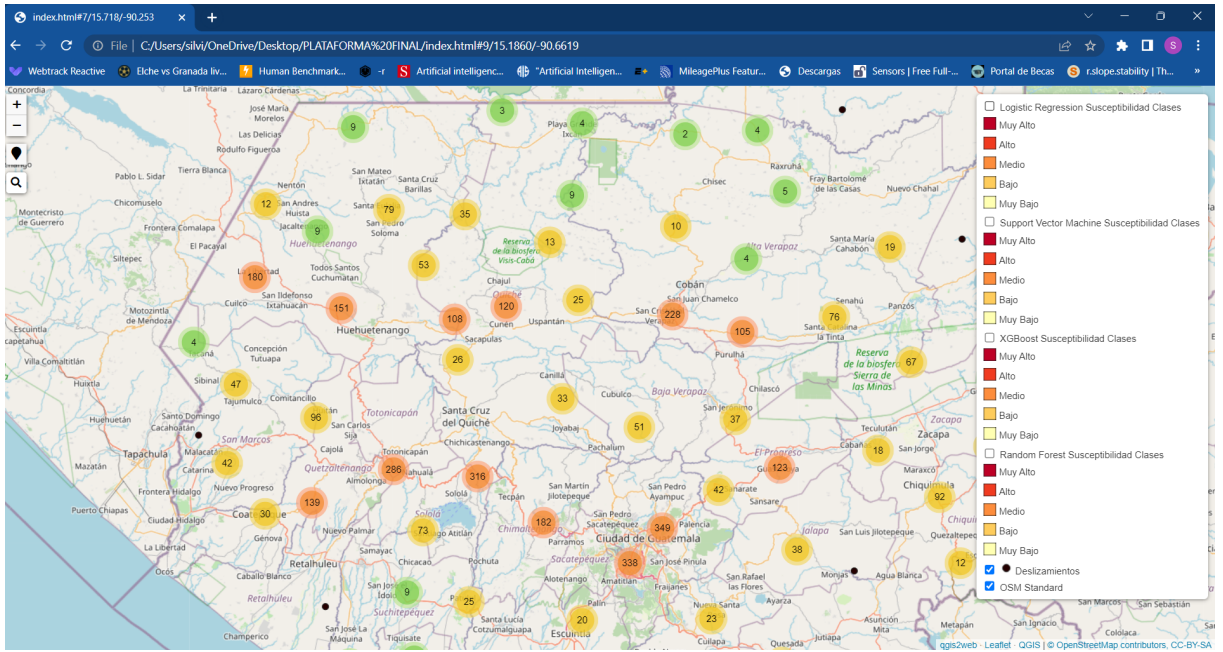
Fuente: elaboración propia

Figura 63: Imagen de plataforma 2



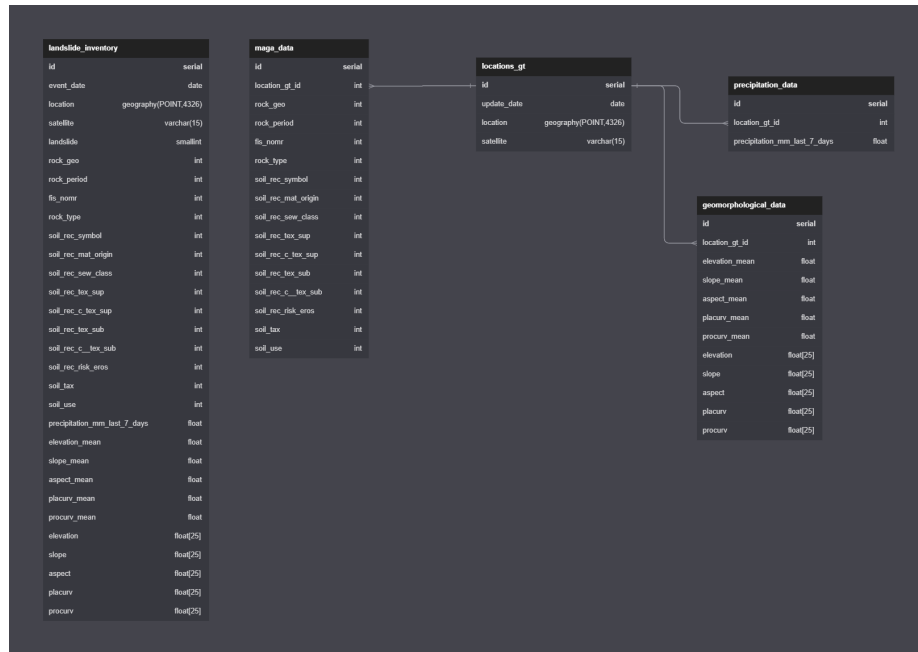
Fuente: elaboración propia

Figura 64: Imagen de plataforma 3



Fuente: elaboración propia

Figura 65: Diseño de la base de datos



Fuente: elaboración propia

6.10. Modelos de segmentación semántica de imágenes

6.10.1. Resultados de los diferentes modelos

Cuadro 82: Resultados de los modelos evaluados por la métrica miou

Modelo	no- <i>data</i>	background	building	road	<i>forest</i>	agriculture	mIoU
U-Net	0.9258	0.6935	0.5614	0.5400	0.5179	0.6605	0.6498
FPN	0.9262	0.6601	0.5498	0.0013	0.5496	0.6676	0.5591
<i>PSPNet</i>	0.9121	0.5979	0.001	0.001	0.2669	0.3946	0.3623
PANet	0.9337	0.6637	0.5480	0.0013	0.5642	0.6748	0.5643
<i>DeepLab V3+</i>	0.9278	0.7071	0.5851	0.5637	0.5776	0.6895	0.6752

Fuente: elaboración propia

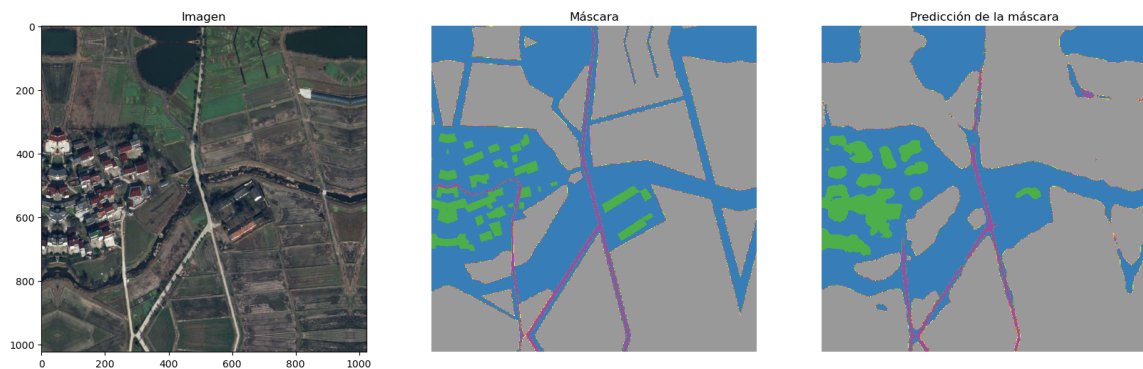
Figura 66: Clases de la segmentación semántica



Fuente: elaboración propia

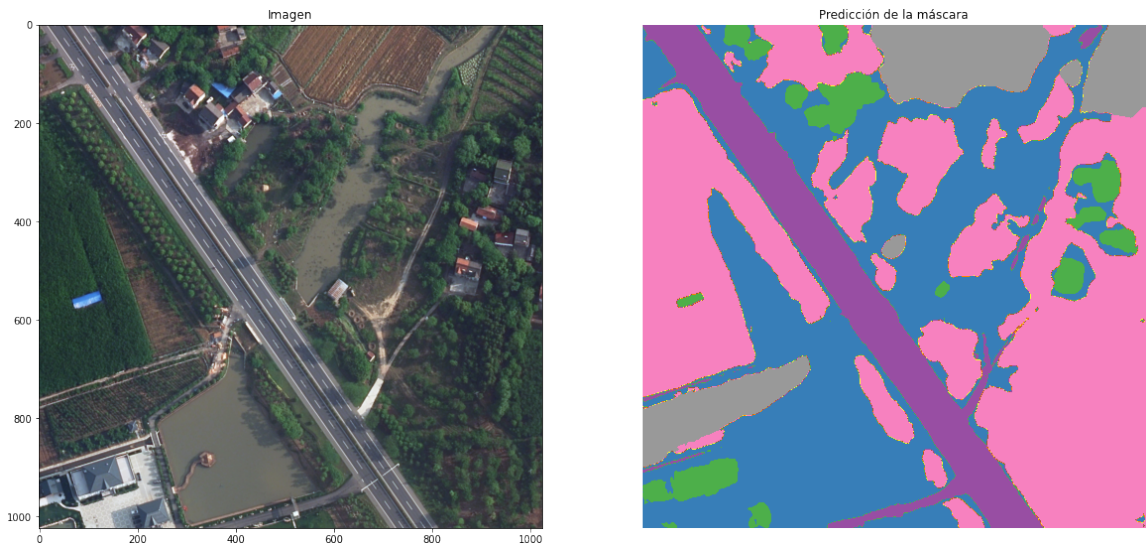
6.10.2. Resultados del modelo final de segmentación semántica

Figura 67: Segmentación semántica con *validation data*



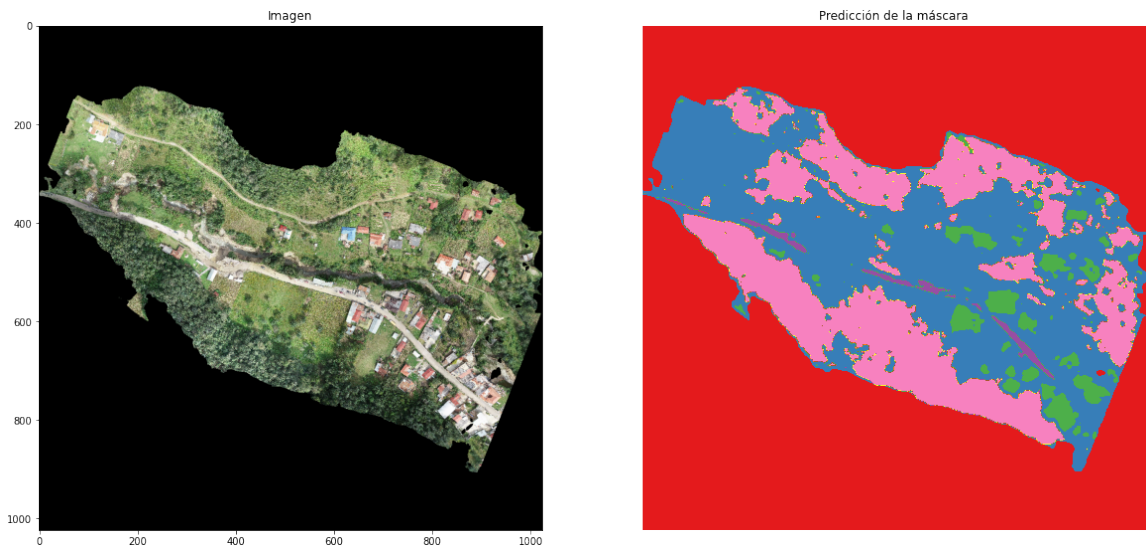
Fuente: elaboración propia

Figura 68: Segmentación semántica con *test data*



Fuente: elaboración propia

Figura 69: Segmentación semántica de ortomosaico



Fuente: elaboración propia

7.1. Sobre la generación del inventario y dataset de deslizamientos

En cuanto a la elaboración del modelo de datos, el inventario de deslizamientos y los *datasets* generados a lo largo del proyecto, para el aprendizaje de modelos de *Machine Learning* que permitieran brindar una susceptibilidad de deslizamientos a nivel nacional en Guatemala. Lo anterior, es de suma importancia y primer paso, en el proceso, dado que un modelo de predicción de susceptibilidad está ligado a las características geomorfológicas, geológicas, litológicas, fisiográficas, topológicas de suelos y mecánicas del suelo, por lo que los modelos de *Machine Learning*, no pueden entrenarse con cualquier dataset, sino que idealmente debe ser de las mismas características y condiciones en las que se utilizará el modelo, por lo que es fundamental tener un dataset propio a las características de Guatemala para conseguir buenos resultados. Como se muestra en los procesos A (Sección 5.3) y B (Sección 5.4) de la metodología, todo parte de la necesidad de acudir a instituciones del Gobierno de Guatemala, que pudiesen aportar datos significativos sobre el estudio de desastres naturales en Guatemala, al igual que estudios y clasificaciones de factores ambientales tales como tipos de suelos, litología, fisiología y geología.

Para la elaboración del Inventario de Deslizamientos se hace necesario acudir a la Coordinadora Nacional de Desastres de Guatemala [CONRED], que lleva un registro histórico sobre los desastres naturales que suceden el país. En teoría, cada vez que sucede un desastre en el país, técnicos y geólogos calificados por esta institución introducen los datos en el Sistema de Manejo de Información en Caso de Emergencia o Desastre [SISMICEDE]. Específicamente, para el caso de deslizamientos, al llegar a evaluar la zona, se ingresa en esta base de datos una nueva entrada de un evento de tipo deslizamiento con las propiedades de la fecha del evento, al igual que la latitud y longitud del punto de derrumbe. Desde este mo-

mento, debemos considerar que el sistema elaborado durante este Megaproyecto ya cuenta con un error leve de factor humano, dado que todos los datos de deslizamientos utilizados son ingresados manualmente por expertos y que, igualmente, el punto de deslizamiento es una estimación, dado que no se puede saber con 100 % de exactitud el punto justo dónde ocurrió el deslizamiento, ni tampoco la hora y fecha precisa del acontecimiento a menos que se haga un estudio a detalle con registros de geoposicionamiento. Sin embargo, por las características de estos desastres, se puede considerar que este primer factor de error es despreciable, tomando en cuenta que los deslizamientos abarcan zonas de varios metros y que la estimación del factor detonante (precipitación) se mide en un área de 2500 metros cuadrados. Simplemente, cabe resaltar que los métodos utilizados hoy en día en Guatemala, para la identificación de Derrumbes, aún son realizados de forma manual y serán grabados únicamente si se despliega el personal de Conred a las áreas de acontecimientos. En el Cuadro 42, se puede observar que gracias a esta base de datos, que contaba con 4230 registros desde el 2008 hasta el 2022, se lograron recuperar 3978 deslizamientos con todas las propiedades a analizar del Cuadro 20.

Seguido del proceso anterior, de obtención de puntos de deslizamientos, se debe discutir sobre el proceso de la generación de puntos de no deslizamiento junto con la generación de todas las propiedades modelo de datos de la Figura 28 necesario para cada uno de los puntos del dataset final. Para la obtención de puntos de no deslizamientos, se puede determinar que se siguió una metodología apoyada en bibliografía existente basada en las correctas asunciones que las áreas y puntos propensos a deslizamientos se encuentran ligados a puntos existentes de un inventario de deslizamientos dada su alta reincidencia (también ligadas a la falta de medidas de mitigación adecuadas que existe a nivel país), de la misma forma que la pendiente es considerada uno de los factores más importantes en la concepción de deslizamientos obteniendo como efecto buscar puntos de no deslizamientos en áreas planas y con bajas elevaciones a partir del uso de herramientas de *QGIS* utilizando correctamente los mapas de inventario y *slope* (ver Figura 42). Al ser este un proceso automatizado, se debe tomar en cuenta que, por las asunciones anteriores, podría existir una muy leve cantidad de los puntos generados que podrían ser propensos a deslizamientos por otras condiciones y factores del terreno. A pesar de ello, esta es una metodología bastante acertada que permite el análisis automatizado y en un futuro, para mejorar este proceso podría ser la validación de expertos en el campo de cada uno de los puntos, teniendo presente que esto conlleva muchos recursos de personas y tiempo. De igual forma se vio una diferencia posicional entre las ubicaciones de no deslizamientos y las ubicaciones de deslizamientos, como se puede observar en las figuras 45 y 46. Esto valida la metodología que se planteó al obtener los puntos en donde no ocurrió un deslizamiento, dado que se tomó en cuenta tanto el mapa de incidencias de deslizamientos, como el mapa de pendiente de Guatemala para obtener estos puntos. El aprobar y validar este paso de la metodología, permitió obtener un total de 5764 puntos de no deslizamientos (ver Cuadro 42) que nutren el dataset de deslizamientos.

Finalmente, sobre la generación del dataset se debe tomar en cuenta las propiedades del modelo de datos que fueron electas como factores importantes en la incidencia de un deslizamiento y las fuentes de datos de donde fueron adquiridas. El área afectada se dividió en 25 celdas de 10x10 metros cuadrados, tomando en cuenta que la celda central era el punto exacto del deslizamiento, con las cuales, se obtuvo los datos geomorfológicos de elevación, pendiente, aspecto y curvaturas del terreno. Además, se obtuvo una media de los factores anteriores del área registrada (25 celdas). Sobre la obtención de estos datos es importante

mencionar que se tomaron de Satélites de entre Mediana y Alta calidad con una resolución de 10 metros y por ello, el tamaño escogido de cada celda. Esta decisión se basa en la forma más rápida, económica y precisa con la que se pueden obtener estos datos, dado que existen otros métodos con mayor resolución como la utilización de Drones (Resolución 3cm), que pueden llegar a obtener mejores resultados, no obstante por tratarse de datos históricos de más 10 años de antigüedad y la escala a analizar (nacional), la utilización de Drones no plantea el acercamiento más adecuado, pero si puede considerarse a futuro como un método de monitorización y de detección automática de derrumbes, complementando los mapas de susceptibilidad generados en este proyecto. Otras tecnologías que pudieron ser utilizadas fueron Satélites de Muy Alta Resolución (Resolución 1 a 3m). El problema con estos es el alto costo y las limitantes, aún hoy en día para Guatemala, de la obtención de estas imágenes, tanto por el costo como el uso destinado de estos satélites enfocados en países europeos y Estados Unidos. Por ello, el método utilizado puede considerarse como válido dado la escala del estudio, la cantidad de datos a utilizar y que se empleó correctamente la herramienta de Google Earth Engine para la obtención de estos datos.

Para observar la resolución de los datos geomorfológicos obtenidos de las imágenes satelitales y compararlos con los de un ortomosaico, se obtuvo y se realizaron mapas de todos estos datos de una misma ubicación en estas dos resoluciones distintas, como se puede ver en la Figura 48. La resolución utilizada fue de 10 m/pix, que es la resolución del satélite sentinel 2, mientras que la resolución del ortomosaico con el que se comparó era de 3 cm/pix. Existen diferencias al momento de tener diferentes escalas, a pesar de que es el mismo terreno el que es analizado. El nivel de detalle que llega a tener cada uno de los datos geomorfológicos en un ortomosaico es muy preciso. Pero a pesar de ello, los datos geomorfológicos de una imagen satelital son bastante similares en apariencia y escala, sobre todo para la elevación, pendiente y aspecto. Mientras que al ilustrar la curvatura del plano y la curvatura de perfil, se puede tener diferencias significativas entre las dos escalas. Esta es una de las limitantes que se encontraron en el desarrollo de este proyecto, dado que la resolución de las imágenes satelitales utilizadas fue la mejor resolución posible de satélites públicos.

La obtención de los datos de fisiografía, geología, uso del suelo y clasificación del suelo fueron obtenidos de otras entidades gubernamentales como el Ministerio de Agricultura y Ganadería [MAGA] redireccionados por otras entidades como Instituto Nacional de Sismología, Vulcanología, Meteorología e Hidrología [INSIVUMEH] y Secretaría de Planificación y Programación de la Presidencia [SEGEPLAN]. Sobre estos datos, cabe destacar que lo mejor para que los modelos sean más efectivos es tener la mayor precisión y descripción de los mismos, al igual que una actualización de entre cada 10 y 20 años de los datos. Se utilizaron mapas que datan del año 2005, por lo que se debe tomar en cuenta la actualización de estos datos en un futuro cercano, de igual manera que es necesario seguir las actualizaciones de estas entidades sobre cualesquiera de los mapas utilizados por el modelo (mencionados al comienzo de este párrafo). La selección de estas variables se fundamentó en los factores permanentes de derrumbes, utilizando el tipo de roca, el período de la roca, el símbolo de reconocimiento del suelo en Guatemala, el material de origen del suelo, la calidad de drenaje del suelo, las texturas superiores e inferiores del suelo (tipo de suelo), el riesgo de erosión del suelo, la clasificación taxonómica del suelo y el uso del suelo, así como la relación del tipo del suelo y sus propiedades mecánicas. Es importante mencionar que la actualización de estos datos debe darse en un período de cada 10 a 15 años, y que siempre será bueno actualizar datos si se cuentan con nuevos estudios de mayor precisión y exactitud. Para esta

actualización se creó un script de python el cual actualiza la base de datos del sistema y entrena nuevamente los modelos. Para el entrenamiento del modelo se dividió la *data* en 70-30 para el *set* de datos de entrenamiento y pruebas respectivamente, como se puede ver en el Cuadro 42. Como factor disparador, únicamente se tomó en cuenta la precipitación de los últimos 7 días, dado que más de un 70 % de los deslizamientos obtenidos tenían como incidencia relacionada la temporada de lluvias en Guatemala. Sin embargo, en el futuro se podría incluir variables como la actividad sísmica, volcánica y el cambio de uso de suelos (como ocurre en las invasiones de terrenos). Con los resultados de los modelos generados se puede confirmar lo valioso de la integración de estos datos a cada uno de los puntos de deslizamiento. Puede discutirse, que a futuro se pueden agregar otros datos disparadores como la ocurrencia de sismos y el factor humano en el cambio del uso de tierra, y factores permanentes como la vegetación. Estos últimos factores mencionados no pudieron ser tomados en cuenta por la escala en la que teníamos las imágenes satelitales, o ya bien porque no existe una medición exacta hoy en día que sea aceptable para el modelo, porque no existen las herramientas y/o mapas necesarios para este cálculo.

Finalmente, como resultado se logra obtener el primer acercamiento reciente de un Inventario de Deslizamientos en Guatemala con un total de 3978 deslizamientos entre 2008 y 2022, que pueden observarse en la Figura 44 y puede ser utilizado para distintas investigaciones sobre estos desastres, al igual que primordialmente para la evaluación de mapas de susceptibilidad y riesgo en Guatemala. A partir de este inventario, se realizó un mapa de incidencias de deslizamientos, el cual se comparó con el mapa de la pendiente de Guatemala obtenido de la NASA, como se puede observar en las figuras 43, 44 y 42. Este mapa de incidencias coincide y se sobrepone al mapa de la pendiente del país y los puntos más fuertes del mapa de incidencias, tienen un valor considerablemente alto de pendiente. Asimismo, los puntos del mapa con una pendiente baja o nula poseen una cantidad significativamente pequeña. Por lo que esto nos indica que hay una relación fuerte y evidente entre la pendiente de un terreno y la susceptibilidad de que ocurra un deslizamiento, según su tipo y propiedades mecánicas. Estos resultados refuerzan y apoyan la teoría, que dentro de los factores más importantes y determinantes que favorecen la ocurrencia de un deslizamiento es la inclinación de los taludes de un terreno.

En resumen, durante todo el proyecto se desarrollaron dos sets de datos: El inventario de deslizamientos y de no deslizamientos y el *set* de datos de los puntos de todo el terreno de Guatemala generados a un kilómetro de distancia entre cada uno. Los dos sets de datos contenían la misma información, dado que tenían la misma estructura del modelo, la cual se puede visualizar en la Figura 47. Las variables que conforman a este modelo fueron escogidas de acuerdo a la investigación previa que se realizó en este proyecto de los factores más importantes e influyentes en deslizamientos de tierra. Más adelante se discutirá la importancia que tuvieron estas variables en los modelos de clasificación que se desarrollaron.

7.2. Sobre los modelos de predicción

Como se mencionó anteriormente, se logró generar un dataset para la creación de modelos de *machine learning* con un total de 9742 puntos de datos, con 3978(41 %) datos con etiqueta de deslizamiento y 5764 (59 %) datos con etiqueta de no deslizamiento. Se decidió dividir el

dataset en dos subgrupos: *train* y *test*, con una proporción de 70:30, según se establece como una de las mejores prácticas para división de un conjunto para el entrenamiento del modelo y un conjunto para la validación del mismo. Por lo tanto, en el conjunto de entrenamiento se tuvo 2760 datos con etiqueta de deslizamiento y 4059 datos con etiqueta de no deslizamiento, teniendo una proporción de 59 % para la etiqueta negativa y un 41 % para la etiqueta positiva. Un dataset ligeramente desbalanceado, por lo que no se procedió a utilizar ninguna técnica de balanceo, dado que se hizo una evaluación de resultados rápida no resultando en gran diferencia y estableciendo que sería mejor conservar los datos únicamente 100 % reales sin datos sintéticamente generados o repetidos. Para el conjunto de *test*, se tuvo 1705 datos con etiqueta de deslizamiento y 1218 datos con etiqueta de no deslizamiento.

Para crear modelos de susceptibilidad de deslizamientos, se debe tomar en cuenta que en Guatemala se ha optado por utilizar métodos tradicionales y basados en que la relevancia es dada por el autor del modelo. Es decir, está sujeta a la subjetividad del experto en cuestión que desarrolló el modelo. Para nuestro caso, se optó por modelos de *machine learning*, que delegan totalmente la responsabilidad al modelo en cuestión sobre las decisiones e importancia que le dan a cada una de las variables predictoras para llevar a cabo una predicción sobre si ocurre un deslizamiento o no. Esto permite darle objetividad al Modelo, únicamente proporcionando al modelo los factores más importantes y determinantes en un deslizamiento según la investigación realizada, pero que sea el modelo quién tome las decisiones de las importancias de las mismas. Por ello, se optó en métodos basados en árboles de decisión; siendo estos 4: *Decision tree*, *Random forest* y *XGBoost*. El primero, siendo el más básico de todos únicamente utilizando un árbol de decisión y los siguientes utilizando métodos de *Ensemble Learning*, distribuyendo a *Random forest* como un método basado en *Bagging* y *XGBoost* métodos basados en *Boosting*. De igual forma, se optó también por utilizar métodos basados en probabilística y estadística, que han sido probados también ser muy buenos por investigaciones pasadas, seleccionando *Logistic regression* y *Support vector machine*. A cada uno de estos modelos, se les ajustó los hiperparámetros (*Hyperparameter Tuning*) utilizando *5 Fold Cross Validation* a través de una cantidad de 20 iteraciones utilizando un *Grid Search* para la mejor selección de parámetros. El paso anterior se realizó empleando únicamente el conjunto de *train* (70 %).

Previo al entrenamiento, se realizó el preprocesamiento de todos los datos. Primeramente, se exploró la *data* que se tenía para conocerla mejor. Luego, de verificar que no existían datos faltantes y nulos, se procedió a verificar las necesidades de cada uno de los algoritmos escogidos para el entrenamiento. Tanto para los métodos basados en árboles de decisión y los métodos basados en probabilística, se procedió a realizar un Label Encoder de todas las variables categóricas, afectando a todos los datos provenientes de los mapas del MAGA, siendo estos los siguientes y utilizados: tipo de roca, período de roca, el tipo de roca, el período de la roca, el símbolo de reconocimiento del suelo en Guatemala, el material de origen del suelo, la calidad de drenaje del suelo, las texturas f e inferiores del suelo (tipo de suelo), el riesgo de erosión del suelo, la clasificación taxonómica del suelo y el uso del suelo. Los datos geomorfológicos que coincidían con los datos numéricos, se decidió realizar una separación de entrenamiento utilizando únicamente la celda 13 (punto exacto del deslizamiento si existió) y una media de las 25 celdas, para realizar una comparación de estos resultados. Para ello, estos datos numéricos siendo elevación, pendiente, aspecto, curvatura del plano y curvatura del perfil, junto con el dato de precipitación de los últimos 7 días, fueron procedidos a utilizar un Standard Scaler (escalar los datos) para los modelos de *Logistic regression* y *Support vector*

machine que lo necesitaban, mientras que para los métodos basados en árboles esto no fue necesario, ya que el modelo no lo requiere. Finalmente, utilizando el lenguaje de Python y las librerías de pandas y SciKIT Learn se procedió a entrenar y validar los 12 modelos realizados: 6 pertenecientes al usar el punto exacto y 6 pertenecientes a utilizar la media de cada una de las variables geomorfológicas.

Para la validación de los modelos, se utilizó el conjunto de *test* y se utilizaron diferentes métricas estadísticas tanto específicas para la evaluación de modelos de aprendizaje supervisado como más exactas para el contexto del problema de deslizamientos. Para la evaluación desde la perspectiva de aprendizaje supervisado se utilizó *ROC AUC* como la principal métrica, junto con *accuracy*, *sensitivity*, *specificity*, y *f1-score*. Por otra parte, para la evaluación desde la perspectiva de deslizamiento se utilizó *Landslide Density (LD)*, *Landslide relative index (R-index)* y *relative error*. Para evaluar *Landslide Density* y *Landslide relative index*, es necesario evaluarlo según las clases de susceptibilidad creadas que pueden identificarse claramente en el Cuadro 55. *Landslide Density* utiliza el conjunto de *test* para la validación, mientras que *Landslide relative index* utiliza el inventario de deslizamientos, para que ambos sean válidos su valor debe aumentar de la clase de susceptibilidad más baja hacia la más alta y la clase de susceptibilidad más alta debería tener valores grandes en comparación de las demás para llegar a aceptar el modelo como confiable. Cada uno de estos, será ahondado sobre cada modelo para finalmente efectuar una aprobación de los modelos seleccionados.

Se realizaron dos entrenamientos de todos los modelos, en el primer entrenamiento se utilizó el punto central del terreno para las variables geomorfológicas (*elevation_13*, *slope_13*, *aspect_13*, *placurv_13* y *procurv_13*), mientras que para el segundo entrenamiento se utilizó la media de los valores geomorfológicos de todo el terreno (*elevation_mean*, *slope_mean*, *aspect_mean*, *placurv_mean* y *procurv_mean*). Luego de cada uno de los entrenamientos de los modelos se obtuvieron las importancias o pesos de todas las variables del *set* de datos. En los cuadros 80 y 81 se pueden observar el resumen de los resultados de validación para cada uno de los modelos utilizando media y utilizando punto exacto respectivamente, con una columna final de aprobación. En las figuras 49 y 50 se pueden observar cada una de las matrices de confusión al momento de realizar la evaluación con el conjunto *test* para cada uno de los modelos entrenados. En cuanto a las métricas, se puede ver una leve mejora, utilizando la media de los factores geomorfológicos contra utilizando únicamente la celda del punto exacto, lo que si puede llegar a cambiar un poco más es la importancia de las variables que se discutirá más adelante. Con base en esos resultados, se decide que la mejor elección es el de la media por tomar más en cuenta la pendiente, reconociendo que este es uno de los factores más relevantes en la incidencia del deslizamiento y que presenta una leve mejora al punto exacto. Ahora se procederá a realizar una comparación general para luego realizar una comparación de cada uno de los modelos, junto con las métricas de evaluación y la importancia que han dado a las variables, al igual que las diferencias entre el modelo al utilizar la media y el punto exacto con un enfoque a quién haya dado mejores resultados, para, finalmente, volver al resumen de los resultados sobre los modelos aceptados.

En general, para los pesos de las variables de todos los modelos que utilizan el punto central del terreno o punto exacto, se puede ver que la variable "*slope_13*" no es muy importante para casi ninguno de los modelos, siendo *random forest* el modelo que más importancia le dio con 4.27%, mientras que *support vector machine* fue el modelo que menos peso le dio con solo un 0.34%. Además de esto, el modelo de regresión logística

le asigno un coeficiente negativo (-19.56 %) a la variable de la pendiente. Esto significa que el modelo relaciona que entre menor sea la pendiente, más probabilidad existe de que haya un deslizamiento. Esta relación no es cierta, ya que se sabe que mientras más empinada es la pendiente, más probabilidad hay de que haya un deslizamiento y que cuando la pendiente es leve o no existe pendiente, no hay riesgo de que haya un deslizamiento.

Por otro lado, con los modelos que no ven solo el punto exacto donde hubo deslizamiento, sino que ven la media de todos los valores geomorfológicos de todo el terreno dividido en 25 celdas de $10 \times 10 m^2$, tuvo resultados más esperados con respecto a la teoría. Los modelos, en su mayoría, le dieron a "slope_mean" más importancia que los modelos anteriores a la variable "slope_13", siendo *random forest* el modelo que más importancia le dio con 6.29 %. Además de esto, el modelo de regresión logística le asigno un coeficiente positivo (10.34 %) a la variable de la pendiente. Esto está acorde con la investigación previa al desarrollo de este proyecto, ya que indica que entre mayor sea la pendiente de un terreno, mayor es la susceptibilidad de un derrumbe. Y además, esto confirma y apoya los resultados del inventario de deslizamiento previamente desarrollado.

La diferencia entre ambos resultados de los entrenamientos posiblemente se deba a que la latitud y longitud de los deslizamientos que se utilizaron para armar el *set* de datos de entrenamiento no eran tan exactas y, por lo tanto, los deslizamientos no ocurrieron justo en el centro del raster generado de datos geomorfológicos. Por lo que probablemente los primeros modelos entrenados utilizaron la pendiente, elevación, aspecto y curvaturas del terreno en una zona en la que no existió ningún deslizamiento. La variable de pendiente es la más afectada con este error, dado que los valores de pendiente de un terreno no son tan estables o iguales para todo el terreno, como lo es la elevación del terreno. Las dos curvaturas del terreno nunca tuvo mayor importancia para ninguno de los modelos entrenados y esto se debe a la limitante de la resolución de las imágenes satelitales utilizadas.

7.2.1. *Decision tree*

Para los modelos de *Decision tree*, podemos decir que las curva de *ROC AUC* es muy general dado la simpleza del modelos, y son quienes tienen el peor resultado con 0.81 para ambos modelos utilizando media y utilizando punto exacto. En la figuras 52 y 54, se puede observar como es que estos modelos han generado su único árbol de decisión. Claramente, se puede distinguir que al ser tan simples, toman el primer elemento que la mayoría de modelos toma en cuenta como uno de los 3 factores más relevantes. Este factor es la elevación y esto tiende a ser porque más del 75 % de los deslizamientos han ocurrido a alturas sobre el nivel del de mar, aproximadamente a más de 800 metros. Esto, aunque pudiese parecer un error, no puede ser interpretado de forma, sino que más bien es congruente con las características del terreno del país sobre las cuales los deslizamientos se desarrollan sobre la parte montañosa del mismo, ya que para que exista un deslizamiento existe un factor clave de inestabilidad por la altura con relación al suelo, dando mayor a paso a un derrumbe, por lo que es claramente aceptable que la propiedad de elevación sea muy importante para todos los modelos. En cuanto al el error relativo, de los modelos de *Decision tree* es aceptable en ambos menor a 0.5, con 0.429 al usar media y 0.383 con punto exacto. Sin embargo, en los cuadros 68 y 74, se puede ver que el factor de LR *index* falla para ambos, ya que existen muchos deslizamientos que se han predecido como susceptibilidad media, teniendo bastante fallo con respecto a

las demás clases para aceptados como válidos. Lo mismo sucede con el *Landslide Density* calculado (ver Cuadro 56) para el modelo que utiliza la media, por el mismo motivo de fallo, teniendo en la susceptibilidad de la clase media. A pesar de que estos modelos, no han sido aceptados por las validaciones respectivas sobre *LD* y *LR Index*, si podemos observar que son un buen primer acercamiento y no se encuentran tan lejos de ser aceptados a pesar de la simpleza del funcionamiento detrás del modelo. En los cuadros 45 y 50 se puede ver que se han dado importancia a las elevaciones mayores al 40 % y a la precipitación cercana al 35 % logrando ver la importancia de estos 2 factores como un acercamiento inicial en la incidencia de deslizamientos en Guatemala.

7.2.2. *Random forest*

Random forest, fue uno de los modelos que obtuvo mejores resultados teniendo una curva *ROC AUC* de 0.95 tanto para el modelo usando la media como el de punto exacto. Se procede a aceptar ambos modelos por las métricas aceptadas y se decide utilizar el de la media por las leves mejoras que presenta en *sensitivity* y *specifity* siendo mejor en ambos, lo cual es importante tanto para la predicción correcta de un derrumbe hacia su importancia de estar en lo correcta de la predicción, como el alertar lo menos posible de falsos positivos. En el Cuadro 57, vemos que la validación de *Landslide Density* es totalmente efectiva, teniendo 6.59 en la clase de susceptibilidad más alta, 1.52, en alta y 0.09 en las más baja. Lo mismo en el Cuadro 69, dónde cada clase de susceptibilidad incrementa el valor del *LR index*. Esto concuerda que el modelo es totalmente aceptable y confiable, junto con las demás métricas de *ROC AUC* en 0.95 y *relative error* en 0.38. También, podemos discutir la clara mejora de utilizar un método de *Ensemble* vs. un solo Árbol de decisión, yendo de 0.81 en *ROC AUC* hacia 0.95 con excelentes resultados. Sobre la importancia de las variables, las podemos observar en el Cuadro 46 la precipitación con un 21 %, la elevación con un 20 %, la pendiente con casi un 7 % y la textura del subsuelo con un 6 % como las más relevantes. Esto nos indica muy buenos resultados para este modelo, y también se decide usar la media en comparación con el punto exacto al ver que se da una mayor importancia a la pendiente que con el punto exacto.

7.2.3. *XGBoost*

XGBoost también obtuvo muy buenos resultados con curvas de *ROC AUC* ambas en un valor de 0.95, al igual que *Random forest*, con la diferencia que este método es basado en *boosting* vs. *Random forest* que se basa en *bagging*. Ambos modelos utilizando la media y utilizando el punto exacto fueron aprobados para este algoritmo, pero de nuevo la media es ligeramente superior, por lo que se decide quedarse con dicho modelo. Además, supera en gran medida en *Relative Error a Random forest* obteniendo un valor 0.3 aceptables para la validación del modelo. En evaluación de *LD Index* en el Cuadro 59 se obtiene un valor de 4.41 para la clase más alta y de 0.12 para la más baja, validando el modelo y *R index* en el Cuadro 71 también se puede observar el incremento de este valor en orden ascendente hacia la clase de susceptibilidad más alta aprobando los resultados. La importancia de las variables es distinta a las de *random forest*, pero tiene muy buenos resultados, ya que diluye un poco más la importancia de las variables sin delegar tanto sobre la elevación y la precipitación. Esto podemos suponer que sucede por la diferencia de estar basado en *boosting*, que intenta

cada vez ir disminuyendo el error respecto al árbol anterior. Las variables más importantes en el Cuadro 47 se puede ver primero el tipo de roca con un 23 %, la precipitación con 9 %, la elevación con 7 %, la taxonomía del suelo con 7 %, el tipo del suelo con un 7 % y el uso del suelo con 6 %, aunque la mayoría de variables tienen un porcentaje de entre 2 y 8 %.

7.2.4. *Logistic regression*

Logistic regression que es diferente a los métodos basados en árboles, también se pudo aprobar en todas las validaciones obteniendo un *ROC AUC* de 0.9, teniendo una menor *accuracy* que los demás con 0.8 y bajando el valor de *sensitivity* y *specifity*. A pesar de no ser el modelo, con mejores resultados, si se puede mencionar que este es uno de los métodos muy parecidos a los tradicionales, ya que se basa en estadística, por lo que es un buen parámetro de comparación con los métodos basados en árboles. Además, es el que tiene mejores resultados en *LD Index*, ya que la clase de susceptibilidad más alta, tiene el valor más alto en comparación con los demás algoritmos, obteniendo un 7.33 para esta clase de susceptibilidad (ver Cuadro 60). La importancia de variables, para este caso, no es medido en porcentaje, sino en el coeficiente de las variables. Se debe recordar que si el coeficiente es positivo, quiere decir que cuando esta variable aumenta, la probabilidad de ser de la clase positiva también aumenta, mientras que si el coeficiente es negativo, cuando la variable aumenta, la probabilidad de ser de la clase positiva disminuye. En el Cuadro 48 se puede ver que los coeficientes de la precipitación, elevación y pendiente son muy altos, por lo que al ser variables numéricas podemos validar claramente que conforme estas variables aumentan, la posibilidad de que exista un deslizamiento también aumenta. Esto último, es relevante dado que confirma la teoría inicial en donde altas precipitaciones inciden como factor disparador, en pendientes muy inclinadas y elevaciones altas.

7.2.5. *Support vector machine*

Support vector machine obtuvo también excelentes resultados, en especial sabiendo que es muy bueno para clasificación binaria. Obtuvo un *ROC AUC* de 0.92 con *accuracy* de 0.85, tanto para el modelo de punto exacto como el modelo usando la media. En el Cuadro 61 se puede visualizar que el valor de *Landslide Density* es muy bueno para cada una de las clases de susceptibilidad y es especialmente alto para la clase muy alta. De igual forma, el valor de *R index* es muy bueno y cumple con todos los requisitos para aceptar el modelo. El error relativo es de 0.452, concluyendo así que este modelo es aceptable y produce resultados adecuados. La importancia de las variables se calcula a través de una permutación en el Cuadro 49, en donde se ve que las 2 variables más importantes son elevación y precipitación acumulada. Luego todas, las demás aportando al modelo en menor medida.

Finalmente, con 4 de los 5 modelos aprobados, se puede aceptar que se han desarrollado 4 modelos distintos adecuados para la susceptibilidad de derrumbes, aceptando *XGBoost*, *random forest*, *support vector machine* y *logistic regression*, siendo este el orden de quien tiene mejores resultados. Gracias a esto se puede observar en las figuras 55 y 56 los mapas de susceptibilidad de deslizamientos en Guatemala desarrollados, entrenados y validados. De igual forma, en la importancia de las variables de los modelos se puede atender que en Guatemala es de suma importancia la precipitación como factor disparador y complementa

la teoría de como el agua satura los suelos y provoca los deslizamientos. Por otra parte, también permite darse cuenta la relevancia del tipo de suelo como factor de deslizamiento, junto a los tipos de roca y período de la misma, al igual que el factor más importante de elevación y pendiente del terreno, que dan como resultado factores condicionantes a la inestabilidad de taludes (ver Figura 51).

7.3. Sobre los mapas de susceptibilidad y la plataforma

Utilizando los modelos aprobados se realizaron mapas de susceptibilidad de toda Guatemala utilizando los puntos generados de Guatemala a 1 km de distancia. De esta manera se clasificó todo el territorio de Guatemala en cinco clases de susceptibilidad a sufrir un derrumbe, las cuales son: Muy baja, Baja, Media, Alta y Muy alta, como se observa en la Figura 55c. Las áreas de alta susceptibilidad que marca el mapa resultante, tienen relación con el mapa de elevación de Guatemala mostrado en la Figura 43 y también tiene relación con el mapa de incidencias generado por el inventario de deslizamientos mostrado en la Figura 42. Estos resultados son los esperados porque las áreas con mayor historial de derrumbes también son las áreas más susceptibles a este fenómeno e igualmente las áreas con la elevación más pronunciada también son las más susceptibles a un deslizamiento.

En conclusión, los mapas de susceptibilidad generados en las figuras 55 y 56, y sus similitudes, nos alientan a los buenos resultados que han obtenido de los modelos para mapear la susceptibilidad de derrumbes en Guatemala, actualizándose constantemente por el factor de precipitación. Al realizar la validación de los mapas generados con ingenieros civiles a través de entrevistas y haber mostrado el sistema, se puede observar la gran importancia de estos mapas para zonificar riesgos de derrumbe en Guatemala, y que estos puedan ser tomados en cuenta para la planificación y construcción futura en Guatemala.

Con lo anterior, se realizó una plataforma la cual muestra los mapas de susceptibilidad de Guatemala generados por los modelos entrenados y también muestra el inventario de deslizamientos recabados en este proyecto, como se puede ver en las figuras 62, 63 y 64. Se incluyó el inventario de deslizamientos para que se pueda realizar comparativas entre este inventario y los mapas de susceptibilidad generados.

Para preservar los datos utilizados por la plataforma y utilizados por el modelo de susceptibilidad de los datos, se guardaron todos los datos generados en una base de datos espacial, y se siguió el diseño que se presenta en la Figura 65. En la base de datos se guardó en dos tablas distintas el inventario de deslizamientos y los puntos generados a lo largo de toda Guatemala con una separación de 1 km entre un punto y otro, y esto debido al periodo de actualización de los datos.

El inventario de deslizamientos se actualiza dentro del sistema dependiendo de la actualización del histórico de derrumbes de la conred, pero se recomienda que estos sean actualizados cada año para que el modelo entrenado tome en cuenta nuevas incidencias de desplazamientos. Mientras que la información de los puntos de Guatemala se actualizan de manera periódica. Los datos geomorfológicos se actualizan cada mes, dado que estos datos pueden cambiar dependiendo de los deslizamientos y sismos que ocurran en este periodo. La precipitación acumulada es un factor disparador, por lo que la actualización de este dato se

realiza cada día. Los datos de los mapas del MAGA e INSIVUMEH son los únicos datos de los puntos de Guatemala que no se actualizan de manera periódica y dependen de nuevas actualizaciones de dichos mapas.

Al tener un sistema que actualiza sus factores de forma automática, permite que los mapas de susceptibilidad cambien y se actualicen dependiendo los cambios en sus datos geomorfológicos y los cambios en sus factores disparadores, que en este caso solo es la precipitación acumulada del terreno. Por lo que en temporada lluviosa las áreas de muy alta susceptibilidad en mapa de susceptibilidad a derrumbes generado debería de ser mayor al área mapeada en épocas secas. Permitiendo todo esto que la plataforma y sus resultados se haga útil para buscar las zonas de riesgo constante para iniciar planes de monitorización específica a mayor detalle, y prevención y mitigación de las zonas más altas (más rojas).

7.4. Sobre los modelos de segmentación semántica

Se realizaron modelos de segmentación semántica para evaluar casos específicos de áreas susceptibles a deslizamientos de tierra y determinar los objetos que se encuentran alrededor de los taludes monitorizados, tales como estructuras, carreteras y elementos en la naturaleza. Eso es útil para realizar un análisis de riesgo de cualquier área, ya que basándose en la cantidad de estructuras, terreno para agricultura, bosque y carreteras o calles, el riesgo de este deslizamiento va a aumentar o disminuir. Dado que no es lo mismo que el deslizamiento ocurra en medio de un barranco a que ocurra al lado de la carretera. Las clases con las que se desarrolló esta segmentación semántica se pueden ver en la Figura 66.

Se desarrollaron 5 modelos de visión por computadora, los cuales fueron: U-Net, *Feature Pyramid Network* (FPN), *Pyramid Scene Parsing Network* (PSPNet), *Path Aggregation Network* (PANet) y *DeepLab V3+*. La métrica que se utilizó para los modelos de la segmentación semántica fue Mean Intersection Over Union (mIoU), dando el valor de IoU para cada clase en todos los modelos, como se puede ver en el Cuadro 82. Los resultados de esta métrica fueron relativamente bajos para todos los modelos. Esto se debe a que se modificaron con la librería *albumentation* tanto los datos de entrenamiento, como los datos de validación. Esta librería permitió que las fotografías que entrenaron al modelo fueran más variadas en tamaño, escala, rotación, brillo y sombras dentro de la imagen. De esta manera se aseguró de que los modelos no realizarán *overfitting* y además que el modelo pudiera tener buenos resultados teniendo alguna o cualquiera de estas variaciones ya mencionadas.

El mejor modelo para la mayoría de las clases fue el modelo de *DeepLab V3+*, ya que tuvo la mayor métrica en las clases background, buiding, road, *forest* y agriculture, con 0.7071, 0.5851, 0.5637, 0.5776 y 0.6895 respectivamente, y solo la clase no-data no tuvo la mejor valoración con 0.9278, siendo el modelo PANet el que tuvo los mejores resultados en esta clase con 0.9337. A pesar de no dominar en todas las clases, *DeepLab V3+* fue el modelo que mayor mIoU obtuvo con 0.6752. El otro modelo que tuvo resultados parecidos aunque con resultados no tan buenos fue el modelo U-Net con un mIoU de 0.6498. Los modelos FPN, *PSPNet* y PANet tuvieron dificultades para predecir la clase road. Esto se puede deber a que esta era la clase que más dificultad presentaba, ya que muchas de las imágenes del *set* de datos tenían máscaras de road con un tamaño significativamente menor al tamaño presentado por las otras máscaras. Por esto mismo se escogió el modelo *DeepLab*

V3+ como el mejor modelo para segmentación semántica.

En la Figura 67 se puede observar el resultado del mejor modelo en datos de validación y se puede comparar la máscara real de la imagen y la predicción de la máscara. El resultado es muy parecido al real, y se puede ver que justamente las clases que mejor puede identificar el modelo es la de agriculture y la de background, mientras que la máscara de road y buiding no se identificaron y tuvieron menos precisión. En las figuras 68 y 69 se obtuvieron máscaras para un dato de prueba y para un ortomosaico. Estas máscaras muestran una precisión muy similar a los resultados de los datos de validación, esto es gracias a las transformaciones que se le hicieron a las imágenes que entreno al modelo.

1. El ángulo de los taludes y la ubicación de los deslizamientos se obtuvieron por medio de visión por computador y geoposicionamiento y fueron fundamentales para el modelo de datos generado.
2. Se utilizaron modelos de segmentación semántica para determinar los objetos que se encuentran en las áreas monitorizadas, los cuales tuvieron muy buenos resultados para identificar y segmentar ortomosaicos e imágenes aéreas.
3. Las variables que conformaban al modelo de datos fueron de utilidad para que los modelos entrenados tuvieran buenos resultados, teniendo como las variables más importantes; la elevación, la pendiente, el tipo de suelo y sus propiedades mecánicas, la precipitación acumulada, el tipo de roca, la calidad de tipo del suelo, el período de roca y el material de origen del suelo.
4. La base de datos espacial desarrollada para la plataforma contribuye a tener un sistema actualizado y garantiza la seguridad e integridad del histórico de desplazamientos.
5. Se generó un inventario de deslizamientos de Guatemala reciente desde el 2008 hasta el 2022, que cuenta con 3978 deslizamientos y varios de sus factores predisponentes a los mismos, esencial para los modelos de susceptibilidad y un sistema de monitorización de derrumbes.
6. Al obtener un dataset con un modelo de datos particular de los factores influyentes de deslizamientos en Guatemala, se logró aprobar y validar 4 modelos de *machine learning* que permiten desarrollar 4 modelos de susceptibilidad a deslizamientos con resultados similares. El modelo con mejores resultados fue *XGBoost*, ligeramente teniendo un *accuracy* de 0.89, *roc auc* de 0.95, un relative error de 0.26 y mostrando excelentes validaciones con *landslide density* y *landslide density r-index*. Se aprueba este modelo como el más confiable y de mejores resultados, seguido de *Random forest*, *Support vector machine* y *Logistic regression*.
7. Los 4 modelos de susceptibilidad generados muestran resultados con un alto nivel de precisión y permiten aprobar que tanto los métodos basados en árboles de decisión

como los métodos basados en estadística y probabilística son buenos acercamientos para la predicción de derrumbes y zonificación de áreas susceptibles.

8. Al poder interpretar las variables que tiene mayor incidencia en los análisis los modelos, se puede llegar a concluir que las más importantes han sido elevación, precipitación acumulada de los últimos 7 días, el tipo del subsuelo y su calidad de textura, la pendiente del terreno y el tipo de roca y su período. Esto concuerda con las investigaciones previas en los factores más importantes para la predisposición de un derrumbe, especialmente tomando en cuenta que en Guatemala estos tienden a suceder en altas elevaciones por sus terrenos montañosos, la precipitación juega un papel muy relevante como disparador de la saturación del suelo y provocación de deslizamientos y la relación de los tipos de suelos y sus propiedades mecánicas correspondientes respecto a su estabilidad.
9. Al integrar el sistema de monitoreo, con todos estos datos de predisposición a deslizamientos, se pueden realizar estudios posteriores que encuentren mayor causalidad entre valores específicos de estos factores y la predicción de probabilidad de que exista un deslizamiento, especialmente con el tipo del suelo, las pendientes y la precipitación acumulada que incita los deslizamientos.
10. Se concluye que la importancia de los modelos de susceptibilidad generados y sus correspondientes, mapas de susceptibilidad, es el poder situarlos a disposición del público y personas expertas en la prevención de desastres naturales, para que puedan ser de gran uso en la prevención, mitigación, gestión y planeación a futuro ante los deslizamientos en Guatemala.

Recomendaciones

1. Se recomienda poder buscar métodos de más alta resolución para que los modelos y sus predicciones sean mucho más exactas y precisas. La generación de inventarios es un proceso largo que implementa y se complementa con diferentes fuentes de información. Sobre los datos geomorfológicos es importante la correcta selección de métodos para obtenerlos y en este proyecto, se optó por la solución más rápida, barata y efectiva, siendo por imágenes satelitales de Google Earth Engine de los satélites Sentinel2. A futuro, puede mejorarse esta recolección utilizando otras imágenes aéreas o satelitales de mayor resolución, pero que pueden llegar a tener un alto costo o bien al necesitar imágenes multitemporales pueden ser una limitante.
2. Se tomaron en cuenta 18 factores predisponentes a los deslizamientos del Cuadro 20. En el futuro, se pueden agregar otros datos como índice de vegetación, cambios de uso de la tierra y mapas de sismología, como otros factores importantes, de los cuales actualmente no se encontraba esta información precisa, por lo que también sería bueno el desarrollo de herramientas que puedan calcular estos factores. Especialmente, el de cambio de uso de la tierra, que implica también uno de los grandes factores de deslizamientos, ingresando las causas y factores humanos.
3. Considerar analizar también la precipitación de próximos 7 días, para poder realizar una predicción de las áreas susceptibles y una comparación de estas áreas con las áreas susceptibles utilizando la precipitación acumulada de los últimos 7 días.
4. Sobre la actualización de los datos en los inventarios de deslizamientos y los mapas utilizados para los factores, se recomienda, como se muestra en el diseño del sistema 61, actualizar los datos de precipitación diariamente, los datos geomorfológicos mensualmente (o encontrando mejores resolución y alternativas) y los demás datos extraídos de mapas cuando se tengan nuevos estudios de mayor calidad y más precisos sobre los mismos de la República de Guatemala.
5. Actualizar el inventario de deslizamientos, preferiblemente cada año, pero dependiendo de las actualizaciones del histórico de derrumbes de la conred, para que el modelo entrenado tome en cuenta nuevas incidencias de desplazamientos recientes.

6. Sobre las zonas más susceptibles a deslizamientos, se recomienda buscar métodos de monitorización más exactos y constantes, el cual se puede realizar con ortomosaicos (modelos de elevación digital generados con drones), ya que este proyecto se basa en una escala nacional y se puede tomar como un primer acercamiento a estas áreas, pero puede complementarse muy bien con otras herramientas y metodologías para la monitorización de posibles derrumbes, apoyándose con estudios más específicos de zonas de alto riesgo y alta incidencia de derrumbes ya sea por falta de medidas de mitigación o la baja calidad en las obras que han buscado reducir la cantidad de derrumbes dadas las condiciones topográficas, fisiológicas, geomorfológicas e industriales del área en cuestión.
7. Dentro del estudio a más detalle se recomienda utilizar los modelos entrenados para segmentación semántica, como el modelo Deeplab v3+, para realizar análisis de riesgo de estas áreas con más detalle.
8. Sobre los proyectos a futuro en este marco, se recomienda que sean multidisciplinarios, incluyendo diferentes expertos en temas de geología, ingeniería civil, ingeniería en computación y antropología, para seguir buscando métodos de mitigación y prevención de derrumbes con innovación de tecnologías en Guatemala.
9. A las instituciones gubernamentales, se les recomienda crear sistema de información geográfica con mayor facilidad de acceso a los mapas de geología, litología y de suelos de Guatemala, al igual que incrementar la información disponible acerca de los parámetros y variables investigadas durante estos estudios.

-
- [1] S. Hergarten, “Landslides,” en *Self-Organized Criticality in Earth Systems*, 1.^a ed. Springer Berlin, Heidelberg, 2002.
- [2] L. M. Highland y P. Bobrowsky, *Manual de derrumbes: Una guía para entender todo sobre los derrumbes*. 1325 del Sistema Geológico de los EUA, 2008.
- [3] E. Sánchez, *Libramiento de Chimaltenango: Derrumbes Continúan afectando la ruta interamericana*, ago. de 2022. dirección: <https://www.prensalibre.com/guatemala/comunitario/libramiento-de-chimaltenango-derrumbes-continuan-afectando-la-ruta-interamericana-breaking/>.
- [4] G. E. Méndez, *¡Rocas gigantes! Derrumbe Bloquea carretera en chiquimula*, jun. de 2022. dirección: <https://www.soy502.com/articulo/rocas-gigantescas-derrumbe-bloquea-carretera-chiquimula-32419>.
- [5] D. Ledezma e I. Fustos, “Interpretaciones geológicas de remociones en masa en latitudes medias: Caso de estudio en la zona de Puerto Saavedra, IX Región.,” Tesis doct., sep. de 2020. DOI: [10.13140/RG.2.2.19482.67522](https://doi.org/10.13140/RG.2.2.19482.67522).
- [6] R. República, *Deslizamientos, hundimientos e inundaciones en 9 departamentos por la lluvia en Guatemala en las últimas horas*, ago. de 2022. dirección: <https://republica.gt/guatemala/deslizamientos-hundimientos-e-inundaciones-en-9-departamentos-por-la-lluvia-en-guatemala-en-las-ultimas-horas-202291618310>.
- [7] Coordinación Nacional para la Reducción de Desastres de Guatemala [CONRED], *SISMICEDA*, jun. de 2022. dirección: <https://conred.gob.gt/sismicede/>.
- [8] B. Thiebes, *Landslide analysis and early warning systems local and regional case study in the Swabian Alb, Germany*. Springer Berlin Heidelberg, 2012.
- [9] S. Mandal y S. Mondal, “Geoinformatics and modelling of landslide susceptibility and risk,” *Environmental Science and Engineering*, 2019. DOI: [10.1007/978-3-030-10495-5](https://doi.org/10.1007/978-3-030-10495-5).
- [10] I. A. Ayala, “Landslides: ideslizamientos o movimientos del terreno: Definicion, clasificaciones y terminologia,” Tesis de mtría., Instituto de Geografía, UNAM, 2000.

- [11] A. Wubalem, "Landslide Inventory, Susceptibility, Hazard and Risk Mapping," en *Landslides*, Y. Zhang y Q. Cheng, eds., Rijeka: IntechOpen, 2021, cap. 11. DOI: [10.5772/intechopen.100504](https://doi.org/10.5772/intechopen.100504). dirección: <https://doi.org/10.5772/intechopen.100504>.
- [12] Secretaría de Planificación y Programación de la Presidencia de Guatemala [SEGEPLAN]. dirección: <https://ideg.segeplan.gob.gt/geoportal/>.
- [13] F. A. Sarría, *Sistemas de Información Geográfica*. Universidad de Murcia, 2006.
- [14] *OGC GeoTIFF Standard*, Open Geospatial Consortium, USA, 2022.
- [15] J. Escalante, H. Porras y J. Cáceres, "Ortomosaicos y modelos digitales de elevación generados a partir de imágenes tomadas con sistemas UAV.," *Tecnura*, págs. 119-140, 2016.
- [16] *ESRI Shapefile Technical Description*, Environmental Systems Research Institute, Inc., Estados Unidos, 1998.
- [17] P. Ramsey y M. Leslie, *Introduction to PostGIS*, PostGIS, Estados Unidos, 2012.
- [18] J. P. Wilson, *Terrain Analysis: Principles and Applications*, J. P. Wilson y J. C. Gallant, eds. Nashville, TN: John Wiley & Sons, jul. de 2000, págs. 51-85.
- [19] C. J. van Westen, E. Castellanos y S. L. Kuriakose, "Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview," *Engineering Geology*, vol. 102, págs. 112-131, 2008.
- [20] Q. Zhou, "Digital Elevation Model and Digital Surface Model," en mar. de 2017. DOI: [10.1002/9781118786352.wbieg0768](https://doi.org/10.1002/9781118786352.wbieg0768).
- [21] Google Earth Engine, *Meet Earth Engine*. dirección: <https://earthengine.google.com>.
- [22] Ministerio de Agricultura, Ganadería y Alimentación, *Atlas temático de la República de Guatemala (Serie de Recursos Naturales, Sociales, Productivos, Amenazas y Vulnerabilidad)*. Unidad de Planificación Geográfica y Gestión de Riesgo (Upggr)., 2005.
- [23] I. Mirsal, *Soil Pollution*. Springer, 2014, ISBN: 9783662054017. dirección: <https://books.google.com.gt/books?id=1Iv-sgEACAAJ>.
- [24] J. Diaz, *Deslizamientos y estabilidad de taludes en zonas tropicales*. Instituto de Investigaciones sobre Erosión y Deslizamientos, Ingeniería de suelos, 1998, ISBN: 9789583308000. dirección: <https://books.google.com.gt/books?id=nWIkAAAACAAJ>.
- [25] K. Kakaes, F. Greenwood, M. Lippincot, P. Meier y S. Wich, *DRONES AND AERIAL OBSERVATION: NEW TECHNOLOGIES FOR PROPERTY RIGHTS, HUMAN RIGHTS, AND GLOBAL DEVELOPMENT*. New America, 2015.
- [26] B. Vergouw, H. Nagel, G. Bondt y B. Custers, "Drone Technology: Types, Payloads, Applications, Frequency Spectrum Issues and Future Developments," Tesis de mtría., Leiden University, 2016.
- [27] K. Ammar, K. Shad, S. Saalem, K. Moinuddin y E. P. Pal, "Comparative Study of Drones: Types and Payloads," Tesis de mtría., Department of Electronics y Telecommunication, 2019.
- [28] T. Jo, *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning*. ene. de 2021, ISBN: 978-3-030-65899-1. DOI: [10.1007/978-3-030-65900-4](https://doi.org/10.1007/978-3-030-65900-4).

- [29] M. Kubat, *An introduction to machine learning*, Third Edition. Springer, 2021.
- [30] J. Korstanje, *Advanced forecasting with python with state-of-the-art-models including LSTMs, Facebook's prophet, and Amazon's DeepAR*. Apress, 2021.
- [31] T. Software, *What is a random forest?* Dirección: <https://www.tibco.com/reference-center/what-is-a-random-forest>.
- [32] T. Zhang, W. Lin, A. Vogelmann et al., "Improving Convection Trigger Functions in Deep Convective Parameterization Schemes Using Machine Learning," *Journal of Advances in Modeling Earth Systems*, vol. 13, mayo de 2021. DOI: [10.1029/2020MS002365](https://doi.org/10.1029/2020MS002365).
- [33] D. G. Kleinbaum y M. Klein, "Introduction to Logistic Regression," en *Logistic Regression: A Self-Learning Text*. New York, NY: Springer New York, 2010, págs. 1-39, ISBN: 978-1-4419-1742-3. DOI: [10.1007/978-1-4419-1742-3_1](https://doi.org/10.1007/978-1-4419-1742-3_1), dirección: https://doi.org/10.1007/978-1-4419-1742-3_1.
- [34] K. Ramasubramanian y A. Singh, *Machine Learning Using R: With Time Series and Industry-Based Use Cases in R*. ene. de 2019, ISBN: 978-1-4842-4214-8. DOI: [10.1007/978-1-4842-4215-5](https://doi.org/10.1007/978-1-4842-4215-5).
- [35] Y. Guo, Y. Liu, T. Georgiou y M. S. Lew, "A review of semantic segmentation using deep neural networks," *International Journal of Multimedia Information Retrieval*, vol. 7, n.º 2, págs. 87-93, nov. de 2017. DOI: [10.1007/s13735-017-0141-z](https://doi.org/10.1007/s13735-017-0141-z), dirección: <https://doi.org/10.1007/s13735-017-0141-z>.
- [36] O. Ronneberger, P. Fischer y T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," en *Lecture Notes in Computer Science*, Springer International Publishing, 2015, págs. 234-241. DOI: [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28), dirección: https://doi.org/10.1007/978-3-319-24574-4_28.
- [37] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan y S. Belongie, *Feature Pyramid Networks for Object Detection*, 2016. DOI: [10.48550/ARXIV.1612.03144](https://arxiv.org/abs/1612.03144), dirección: <https://arxiv.org/abs/1612.03144>.
- [38] H. Zhao, J. Shi, X. Qi, X. Wang y J. Jia, *Pyramid Scene Parsing Network*, 2016. DOI: [10.48550/ARXIV.1612.01105](https://arxiv.org/abs/1612.01105), dirección: <https://arxiv.org/abs/1612.01105>.
- [39] S. Liu, L. Qi, H. Qin, J. Shi y J. Jia, *Path Aggregation Network for Instance Segmentation*, 2018. DOI: [10.48550/ARXIV.1803.01534](https://arxiv.org/abs/1803.01534), dirección: <https://arxiv.org/abs/1803.01534>.
- [40] L.-C. Chen, G. Papandreou, F. Schroff y H. Adam, *Rethinking Atrous Convolution for Semantic Image Segmentation*, 2017. DOI: [10.48550/ARXIV.1706.05587](https://arxiv.org/abs/1706.05587), dirección: <https://arxiv.org/abs/1706.05587>.
- [41] J. Wang, Z. Zheng, A. Ma, X. Lu e Y. Zhong, "LoveDA: A Remote Sensing Land-Cover Dataset for Domain Adaptive Semantic Segmentation," oct. de 2021.

