ASSESSMENT OF SUSCEPTIBILITY TO LANDSLIDE THROUGH DEEP LEARNING ALGORITHMS

A PROJECT REPORT

Submitted by

SHIYAM SUNDAR T	19106108
SHOBIA D	19106109
SOURNAMALYA S B	19106113
THILIBAN R	19106126
VENKATESHWARAN M	19106130

in partial fulfilment for the award of the degree

of

BACHELOR OF ENGINEERING

in

ELECTRONICS AND COMMUNICATION ENGINEERING

HINDUSTHAN COLLEGE OF ENGINEERING AND TECHNOLOGY

Approved by AICTE, New Delhi, Accredited with 'A' Grade by NAAC

(An Autonomous Institution, Affiliated to Anna University, Chennai)

COIMBATORE – 641 032

APRIL 2023

HINDUSTHAN COLLEGE OF ENGINEERING AND TECHNOLOGY

Approved by AICTE, New Delhi, Accredited with ‘A’ Grade by NAAC
(An Autonomous Institution, Affiliated to Anna University, Chennai)
COIMBATORE – 641 032

BONAFIDE CERTIFICATE

Certified that project report “**ASSESSMENT OF SUSCEPTABILITY TO LANDSLIDES THROUGH DEEP LEARNING ALGORITHMS”** is the bonafide work of **SHIYAM SUNDAR T (19106108), SHOBIA D (19106109), SOURNAMALYA S B (19106113), THILIBAN R (19106126)** and **VENKATESHWARAN M (19106130)** who carried out the project work under my supervision.

SIGNATURE	**SIGNATURE**
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Dr.P.Vijayalakshmi, M.E., Ph.D.	**Dr.J.Ramya, M.E., Ph.D.**
Professor and Head of Department,	**Associate Professor ,**
Department of Electronics and	Department of Electronics and
Communication Engineering.	Communication Engineering.
Hindusthan College of Engineering	Hindusthan College of Engineering
and Technology, Coimbatore-641032	and Technology, Coimbatore-641032

Submitted for Autonomous Project Viva-Voice examination held on ………………

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

We would like to express our profound gratitude to our Honourable Chairman, **Shri. T.S.R. KHANNAIYANN** and our Managing Trustee, **Smt.T.R.K. SARASUWATHI** of Hindusthan Educational and Charitable Trust for providing the necessary facilities and support for successful completion of the project within the college.

We extend our sincere thanks to our Chief Executive Officer, Hindusthan Institutions **Dr. K. KARUNAKARAN, Ph.D.,** for his constant support and motivation.

We would like to express our heartfelt thanks to our esteemed Principal, **Dr. J. JAYA, M.Tech., Ph.D.,** for her constant motivation and encouragement.

We wish to express our sincere thanks and gratitude to our Head of the Department **Dr. P. VIJAYALAKSHIMI M.E, Ph.D.,** for her encouragement, guidance and support to complete the project.

We are highly indebted to our Project Guide **Dr. J . RAMYA, M.E, Ph.D.,** for her valuable guidance and supervision in completing the project.

We extend our special gratitude and thanks to our Project Coordinator **Dr .P.K . POONGUZHALI, M.E., Ph.D.,** for her valuable guidance and support to complete the assigned task.

We are also thankful to all our teaching and non-teaching staff members of Department of Electronics and Communication Engineering for their kind cooperation and encouragement

ABSTRACT

Landslide susceptibility refers to the likelihood or probability of an area experiencing landslides. It is determined by analyzing various factors such as slope angle, soil type, land use, vegetation cover, geological structure, and precipitation patterns. Landslide susceptibility mapping is a useful tool for identifying areas that are more susceptible to landslides, which can help in land-use planning and management, hazard assessment, and disaster risk reduction. This research introduces a novel method for landslide prediction utilizing machine learning methods, notably Convolutional Neural Network (CNN) and Support Vector Machine (SVM). Early detection can prevent fatalities and reducedamage from landslides, a natural calamity that can seriously harm infrastructure and result in fatalities. The CNN and SVM algorithms in the proposed model, trained using a set of topographical data from satellite. The SVM is used for classification, whereas the CNN is utilized to extract features. The model's performance in predicting landslides was highly accurate when tested against a dataset of landslides that occurred and did not occur. The findings imply that the suggested method is a promising strategy for early landslide prediction and can be incorporated into current disaster management systems to speed up reaction times and lessen the impact of landslides.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Network
AUC	-	Area under the curve
CNN	-	Convolutional neural networks
DEM	-	Digital elevation model
DT	-	Decision tree
DTM	-	Digital terrain model
LR	-	Logistic regression
RF	-	Random forest
ROC	-	Receiver operating characteristic
SVM	-	Support Vector Machines

CHAPTER 1

INTRODUCTION

In many regions of the world, landslides are a common natural occurrencethat can have disastrous effects. Following an earthquake or a period of intenserainfall, there are thousands of small- to medium-sized ground movements. Climate change, population increase, and haphazard urbanization in unstable mountainous regions have all contributed to landslides becoming increasingly destructive in recent years. Early landslide detection is essential for effective reaction and consequence management.

A landslide's intensity can range greatly, from modest soil and rock movements to catastrophic occurrences that can result in extensive damage and fatalities. Rock falls, debris flows, and earthflows are three different types of slides that may be categorized depending on their features. Rock falls happenwhen isolated boulders or stones fall from cliffs or other steep slopes. On the other hand, debris flows consist of a mixture of dirt, rock, and water that flow down a slope like a fluid. Earthflows, on the other hand, entail the gradual descent of saturated soil and silt. In addition to the immediate harm a landslide causes, the material it leaves behind can obstruct highways and rivers, causing more harm and disturbance. It might be difficult to prevent landslides, but steps like slope stability, drainage, and vegetation control can help lower the danger.

The effects of landslides when they do occur can also be lessened with the use of early warning systems and evacuation preparations. Across the world, landslides pose a serious hazard to property, infrastructure, and human life. To lower the danger of catastrophic events and lessen their impact, accurate landslide forecasting is essential. Manual inspection and observational approaches are significantly used in conventional methods like slope stability studies and geotechnical surveys. To deliver more precise landslide predictions, predictive models can now be trained on enormous amounts of data thanks to improvements in machine learning techniques. In this study, we suggest using Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) to predict landslides based on historical data. Enhancing the timeliness and accuracy of landslides is the goal of this initiative.

1.1 OBJECTIVE

The objective of the landslide prediction project using CNN and SVM is todevelop an accurate and efficient model that can predict the occurrence of landslides based on various environmental factors. The project aims to utilize convolutional neural networks (CNN) to extract features from the input data, and support vector machines (SVM) to classify the data and make predictions. The primary goal is to create a reliable system that can be used by authorities to issue early warnings and take necessary precautions to prevent potential disasters caused by landslides.

By combining these two models, we can develop a more robust and accurate system for landslide prediction that takes into account both the spatial features and the terrain characteristics of the area under study. The ultimate objective ofthis system is to provide early warning to people living in areas prone to landslides, which can help prevent loss of life and property damage

1.2 PROBLEM STATEMENT

Landslide prediction is a critical task in geology and civil engineering. Theoccurrence of landslides can cause severe damage to infrastructure and loss ofhuman life. In recent years, there has been a growing interest in utilizing machine learning techniques to predict landslides. In this project, we aim to utilize convolutional neural networks (CNN) and support vector machines (SVM) for landslide prediction

The dataset for this project consists of satellite images of areas prone to landslides. Each image is labeled as either "landslide" or "non-landslide". We will use the CNN model to extract relevant features from the images, and the SVM model to classify them.

Our objective is to evaluate the performance of these models in terms of accuracy, precision, recall, and F1-score. We will also compare the performance of CNN and SVM models and identify which one performs betterfor landslide prediction. The ultimate goal of this project is to provide an accurate and efficient tool for predicting landslides that can be used by geologists and civil engineers to prevent damage and loss of life caused by landslides

CHAPTER 2

LITERATURE SURVEY

Ding, A, Zhang et al CNN-based automatic landslide recognition and texture change detection The protection of human life and property is seriously threatened by landslides, a common calamity. Remote sensing technology has increased our ability to detect disasters more effectively. Traditional methods, on the other hand, frequently require human involvement, wasting both human labor and material resources. This research proposes a unique approach for automatic landslide recognition based on texture change detection and convolutional neural network (CNN)[1]. Using these processes, which outperform the conventional methods in terms of avoiding a significant amount of mistake detection, efficiency, and convenience, gradually reduces the search scope and then confirms the actual landslide areas.

Haojie Wang et al Employing object-oriented techniques to describe the spectral, spatial, andmorphometric characteristics of landslides for semi-automatic detection An essential component of pre- and post-disaster hazard analysis is the identification and classification of landslides. This has mainly been accomplished through human picture interpretation or field mapping. The spectral, spatial, and morphometric characteristics of landslides can be used to create an object-based classification technique, which can then be used to analyze images semi-automatically[2]. This is a challenging undertaking because recent landslides do not have distinctive shapes and their spectral properties are almost comparable to those of other natural objects, such river sand and rocky outcrops. This study explores the use of spectral, shape, and contextual data in conjunction to identify landslides. The method is put to the test on a region in India's craggy Himalayas using 5.8 m multispectral data from Resourcesat-1 and a 10 m digital terrain model created from 2.5 m Cartosat-1 imagery. For object-oriented analysis, it leverages objects that are created as a result of segmenting a multispectral image as classification units. In the beginning, spectral data along with shape and morphometric parameters were employed to distinguish landslides from false positives. Debris slides were the next classification for items identified as landslides based on the type of material and movement. Using morphometric criteria and adjacency, debris flows and rockslides are studied. They were further categorized based on the terrain curvature failure mechanism. The approach was created for a training catchment and subsequently utilized on an independent catchment without additional modification. With 76.4% recognition and 69.1% classification accuracy, our technique was able to identify five different types of landslides. This technique has the potential to assist risk analysis, disaster management, and decision-making procedures in the wake of an earthquake or an extreme rainfall event because it can identify landslides relatively fast.

Pradhan, Kim et al study over Vulnerable geographical position In order to anticipate landslides using machine learning methods, carried out a case study in Nepal [3]. Data on landslides and numerous environmental elements, such as rainfall, soil type, slope, elevation, and vegetation cover, were gathered by the authors. To create landslide prediction models, they next employed machine learning algorithms like logistic regression, decision trees, and random forests. The random forest method beat the other models in terms of predicting landslides, according to the authors' accuracy metrics evaluation of the models' performance. The study emphasizes the value of applying machine learning approaches for landslide prediction and risk assessment, particularly in areas like Nepal that are vulnerable to landslides

S.N Selamat , N.A.Majid et al Landslide susceptibility assessment using machine learning algorithms The focus of the Kassim et al. (2021) research, "Landslide susceptibility assessment using machine learning algorithms: A case study of Kelantan, Malaysia," is on the application of machine learning algorithms to evaluate landslide susceptibility in the Kelantan area of Malaysia [4] The study used three machine learning models—logistic regression (LR), support vector machine (SVM), and random forest (RF)—to build a map of the area's vulnerability to landslides based on a total of 14 landslide-related parameters, including slope, aspect, curvature, geology, and land use. Receiver operating characteristic (ROC) curves and area under the curve (AUC) values were used to assess the performance of the models. The findings revealed that the RF model, which had an AUC value of 0.945, performed the best, followed by the SVM model, which had an AUC value of 0.938, and the LR model, which had an AUC value of 0.928. Slope, land use, soil texture, and drainage density were among the study's findings about the most significant elements influencing vulnerability to landslides.

In conclusion, the work by Kassim et al. (2021) shows the efficiency of machine learning algorithms in landslide susceptibility mapping and offers useful data for disaster management and land-use planning in the Malaysian province of Kelantan. The results of this study can also be applied to other areas with comparable topographical and geological features.

Tapas R. Martha, Norman Kerle,et al comparison of machine learning algorithms for landslide susceptibility mapping: a case study in the LangSon city area, Vietnam The goal of the paper by Luong et al. (2020) titled "A comparison of machine learning algorithms for landslide susceptibility mapping: a case study in the Lang Son city area, Vietnam" is to assess and contrast the effectiveness of different machine learning algorithms in predicting landslide susceptibility in the Lang Son city area of Vietnam. Five machine learning models—the decision tree (DT), random forest (RF), support vector machine (SVM), artificial neural network (ANN), and logistic regression (LR) models—were used to create a map of the region's susceptibility to landslides based on a total of 11 landslide-related factors, including slope, aspect, curvature, lithology, and land cover. Receiver operating characteristic (ROC) curves and area under the curve (AUC) measurements were used to assess each model's performance. The outcomes demonstrated that the RF model, which had an AUC value of 0.919, performed the best, followed by the ANN model, which had an AUC value of 0.904, the SVM model, which had an AUC value of 0.898, the DT model, which had an AUC value of 0.882, and the LR model, which had an AUC value of 0.867. Slope, aspect, curvature, and lithology were shown to be the most significant determinants of landslide vulnerability in the study.

In conclusion, the work by Luong et al. (2020) shows the efficiency of machine learning algorithms in landslide susceptibility mapping and offers useful data for disaster management and land-use planning in the Vietnam region around Lang Son city [5]. The findings of this study can also be extended to other regions with similar topographical characteristics.

Tran, phong & et al Mapping Landslide Susceptibility Using Machine Learning Algorithms and GIS. In their paper titled "Mapping Landslide Susceptibility Using Machine Learning Algorithms and GIS: A Case Study in Shexian County, Anhui Province, China," Wang et al. (2020) discuss how to use GIS (geographic information system) and machine learning algorithms to map landslide susceptibility in the Chinese county of Shexian.The study used three machine learning models—the random forest (RF), support vector machine (SVM), and artificial neural network (ANN) models—to generate a map of the area's susceptibility to landslides based on a total of 11 landslide-related factors, including elevation, slope, aspect, curvature, lithology, and land use. Receiver operating characteristic (ROC) curves and area under the curve (AUC) measurements were used to assess each model's performance. The findings revealed that the RF model, which had an AUC value of 0.949, performed the best, followed by the ANN model, which had an AUC value of 0.932, and the SVM model, which had an AUC value of 0.917. The Wang et al. (2020) study illustrates the efficiency of machine learning algorithms and GIS in landslide susceptibility mapping and offers useful data for disaster management and land-use planning in the Shexian County of Anhui Province, China [6]

Wang et al Artificial Neural Network An approach for identifying landslides using machine learning strategies is presented in the publication "Landslide identification using machine learning" by Wang et al. (2021).The investigation was carried out in a region of China's Three Gorges Reservoir noted for its intricate terrain and geology and vulnerable to frequent landslides. The study used remote sensing information from satellites like

Landsat-8 and Sentinel-1 to extract several landscape parameters including elevation, slope, aspect, curvature, and vegetation indices. Following that, several machine learning models, such as the Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network were trained using the retrieved features. (ANN). An approach for identifying landslides using machine learning strategies is presented in the publication "Landslide identification using machine learning" by Wang et al. (2021). The investigation was carried out in a region of China's Three Gorges Reservoir noted for its intricate terrain and geology and vulnerable to frequent landslides. The study used remote sensing information from satellites like Landsat-8 and Sentinel-1 to extract several landscape parameters including elevation, slope, aspect, curvature, and vegetation indices. Following that, a number of machine learning models, such as the Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network were trained using the retrieved features. (ANN). Using accuracy criteria including overall accuracy, kappa coefficient, and receiver operating characteristic (ROC) curve, the models' performance was assessed. With an overall accuracy of 94.48%, the findings demonstrated that the RF model performed better than the SVM and ANN models. The results of this study show how machine learning techniques may be used to identify landslides using remote sensing data. The suggested approach may be used to locate probable landslide hotspots and give useful data for disaster management and prevention in other places. The study also underlines the significance of topographical characteristics in landslide identification and the demand for precise topographical information for successful landslide prediction and prevention.

CHAPTER 3

EXISTING METHOD

Landslide identification plays an important role in landslide risk assessment and management, With the advent of the remote sensing technology,landslides can be identified through visual interpretation of both remote sensing images and topographic surfaces. Existing system on landslide identification aremainly based on optical images using pixel-based or object-oriented methods, and the digital terrain model (DTM) is often used as auxiliary data for such analysis.

Combined optical images and digital elevation model (DEM) derivatives to identify translational landslide scars using object-oriented methods. Althoughthe visual interpretation has high identification accuracy, the process is time consuming and labor-intensive. The existing system mainly focuses on identifying landslides with available optical images. The identification of relict landslides is barely explored.

The existing system pays more attention to the identification process itself. The existing system relies heavily on the optical remote sensing imagesto conduct landslide identification. The latest high-resolution DTM can preciselycapture minor terrain differences; yet the potential of DTM-dominant landslide identification using machine learning and deep learning has not been exploited

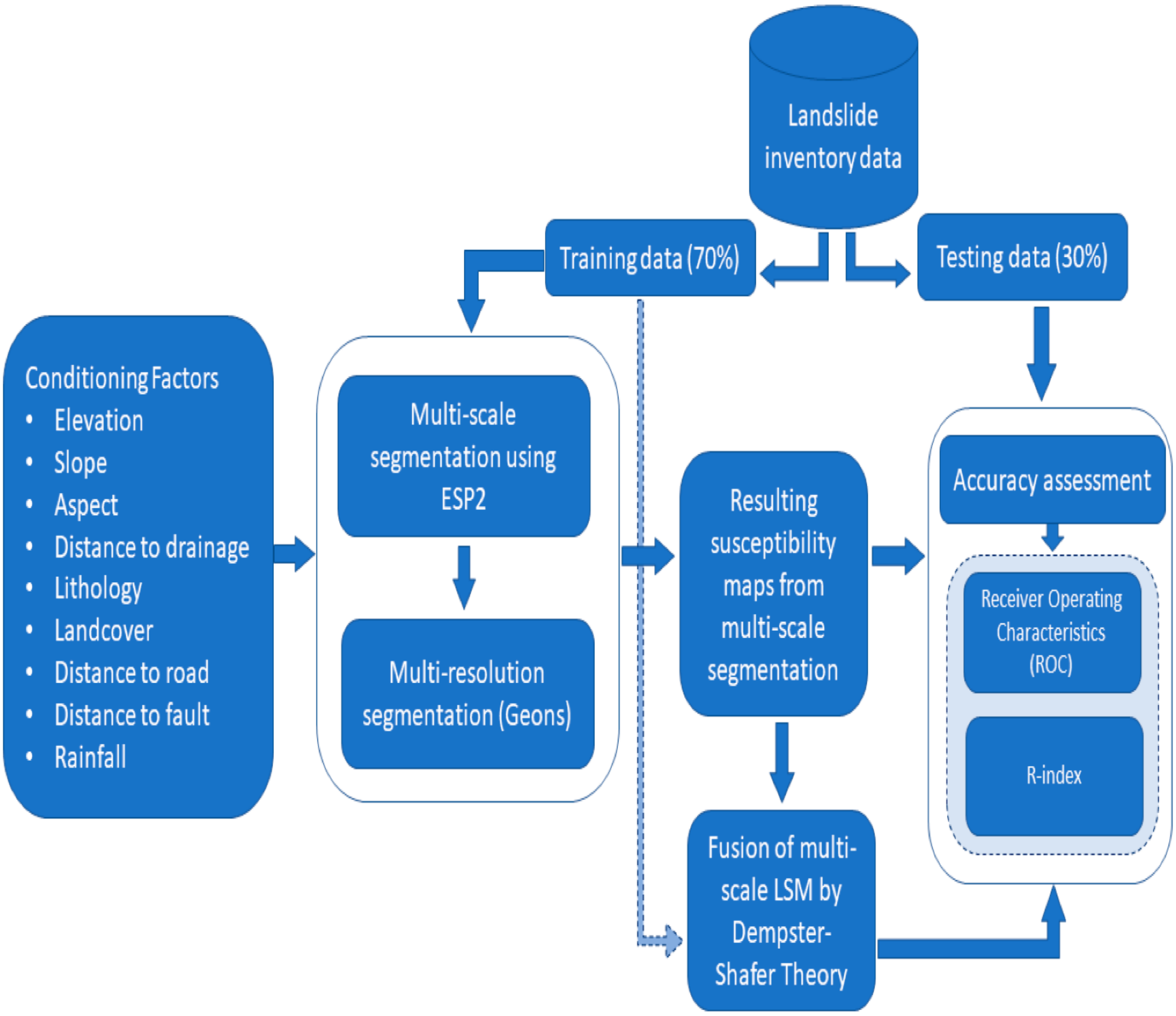


Figure 3.1: Existing methodology block diagram

DISADVANTAGES

While satellite images can provide valuable data for assessing landslide susceptibility, there are also some disadvantages to their use. Some of these disadvantages include:

Limited resolution: Satellite images may not have sufficient resolution to accurately identify small-scale features that contribute to landslide susceptibility

Dependence on weather conditions: Weather conditions can impact the quality of satellite images, and cloud cover or poor lighting conditions can limit the amount of useful data that can be obtained.

Incomplete coverage: Satellite images may not cover all areas of interest, and some areas may be obscured or inaccessible due to vegetation cover or topography.

Cost: The cost of obtaining and processing satellite images can be high, especially for high-resolution images, which may be necessary for accurate landslide susceptibility assessments.

Interpretation challenges: Interpreting satellite images requires specialized knowledge and skills, and even with expert interpretation, there may be uncertainties and inaccuracies in the resulting landslide susceptibility maps.

Difficulty in identifying some landslide triggers: While satellite images can identify some potential triggers of landslides such as changes in land cover, it may not be able to detect triggers such as pore water pressure build up in the soil, which requires more ground-based data collection.

CHAPTER 4

PROPOSED METHOD

4.1 OVERVIEW OF METHOD

Convolutional neural networks (CNN) are potent deep learning algorithms frequently utilized for feature extraction and image processing. Support Vector Machines (SVM), on the other hand, are frequently employed for classification tasks because they can handle complex data and produce high accuracy. A more reliable and accurate model for landslide prediction can be produced by merging CNN with SVM, taking advantage of the advantages of both methods. CNNs can extract detailed elements from satellite photos, such as textures, forms, and patterns, which can be difficult to detect using conventional techniques. After that, the SVM can successfully categorize and forecast the likelihood of landslides using these features as input. Moreover, combining CNN and SVM helps lessen the chance of overfitting, which occurs when the model grows overly complex and starts to perform badly on new data. The SVM may provide a more generalizable model by choosing only the most pertinent characteristics, whereas the CNN can learn the crucial features from the data.

4.2 TRAINING THE DATASET

The quality and dependability of the model must be ensured by a number of processes during the training of a dataset for landslide prediction utilizing Convolutional Neural Network (CNN) and Support Vector Machine (SVM).The first stage is to compile a dataset of satellite pictures and topographical information that includes both actual and hypothetical landslides. For the model to be trained on a representative sample, the dataset needs to be diverse, including many locations and types of terrain. Then, the data must be preprocessed and translated into a format appropriate for usage with the CNN and SVM algorithms. To avoid causing the model to become sidetracked by unimportant details, the satellite images should be scaled, normalized, and any noise or artefacts removed. To guarantee that the topographical data meets the model's input specifications, it should also be preprocessed. Following preparation, the data is divided into a training set and a validation set. The validation set serves as a check for the model's accuracy and a safeguard against overfitting while the training set is used to train the model. CNN is initially trained on satellite photos, utilizing the training set to discover the pertinent characteristics of the data. The SVM, trained on the topographical data, uses these features as input to determine the likelihood of a landslide occurring. These features are then extracted. The model's performance is continuously assessed on the validation set during training, and modifications are made to maximize the model's accuracy. Until the model performs satisfactorily on the validation set, this step is repeated. In conclusion, training a dataset for landslide prediction using CNN and SVM entails prepping the data, dividing it into training and validation sets, training the CNN on satellite images, training the SVM on topographical data, and continuously assessing and optimizing the model's performance

4.3 TESTING THE DATASET

Testing the dataset for landslide prediction using CNN and SVM involves evaluating the performance of the trained models on a separate set of data that was not used during the training process. Load the trained models: The first step in testing the dataset is to load the trained models, which were developed using the training set. For CNN, this involves loading the weights and biases of the trained model, while for SVM, this involves loading the learned parameters. Prepare the testing dataset: The testing dataset should be separate from the training and validation datasets and should include data that the models have not seen before. This testing dataset can be used to gauge how well the models perform with fresh, untested data.

Prediction: The trained models will be used to make predictions on the testing dataset in the following stage. For CNN, this entails running the training dataset through the model before getting the anticipated results. In the case of SVM, this entails predicting the class labels of the testing dataset using the learnt parameters.

Evaluation of performance: The trained models' performance on the testing dataset is assessed as the last stage. Many performance indicators, including accuracy, precision, recall, F1 score, and confusion matrix, can be used to accomplish this. These indicators can show how effectively the models are working and can show where they need to be improved. Overall, loading the trained models, getting the testing dataset ready, making predictions, and assessing the performance of the models are all steps in the process of utilizing CNN and SVM to test the dataset for landslide prediction. The ability of the trained models to perform effectively and generalize well about new and untried data both depend on proper testing.

4.4 ARCHITECTURE

Data cleaning, transformation, and normalization are the initial steps in the preparation of data. Techniques like feature engineering, data normalization, and standardization may be used for this After preprocessing the data, the CNN architecture is trained using the data. Several pooling layers, fully connected layers, and convolutional layers can all be found in the CNN design. In order to extract features from the input pictures, convolutional layers are utilized. Pooling layers are then used to minimize the dimensionality of the feature maps. The input photos are categorized as landslides or non-landslides using the fully linked layers. When the CNN architecture has been trained, the output of the last layer is sent into the SVM classifier. Landslide and non-landslide photos are divided into two groups using the SVM classifier during training. A kernel function, a decision boundary, and support vectors can make up the SVM architecture. Combining the CNN and SVM classifier outputs using an ensemble approach is the final step. To get a final prediction, this might incorporate methods like averaging or weighted averaging

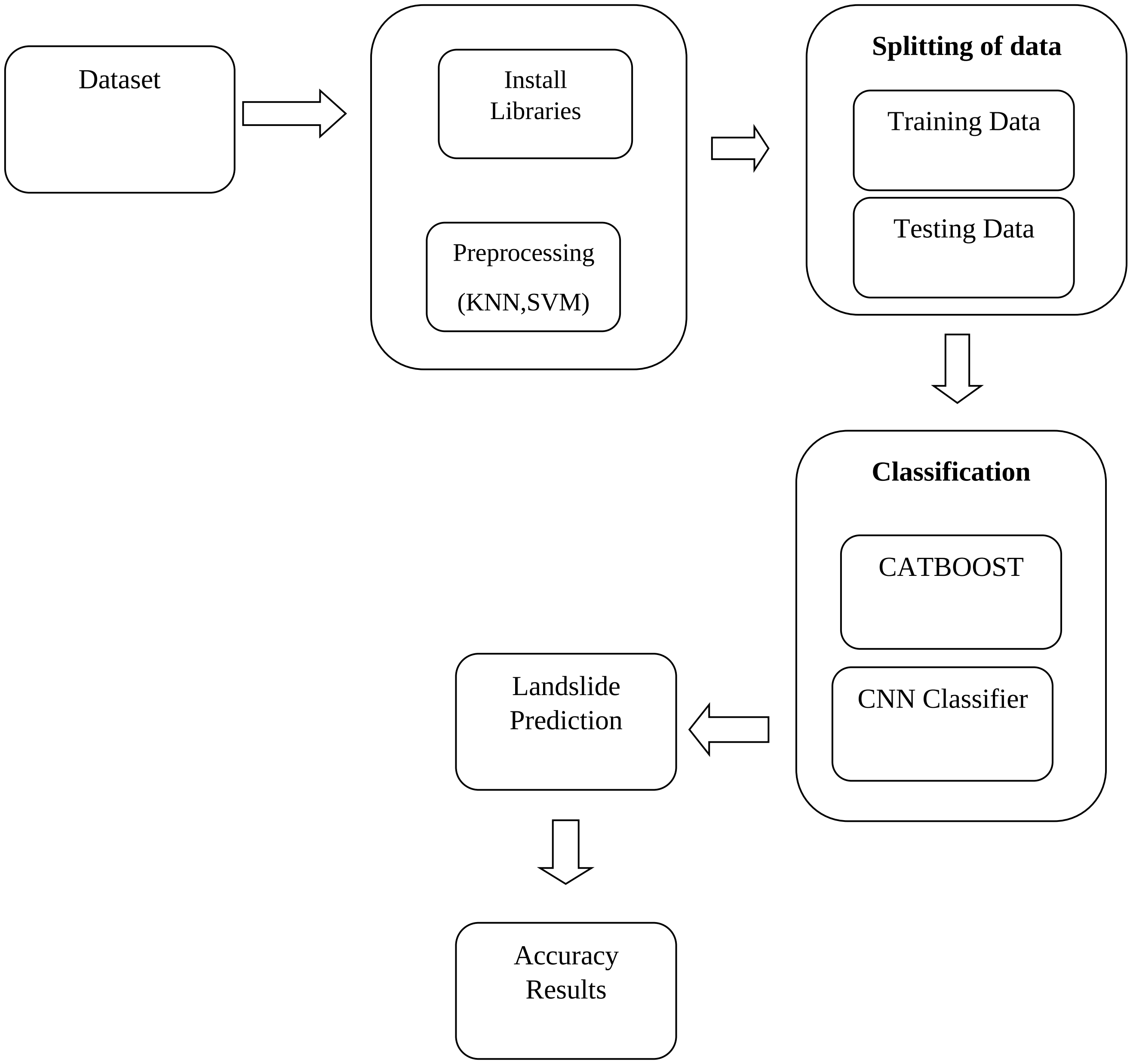


Figure 4.1: Architecture diagram

CHAPTER 5

METHODOLOGY

The methodology we intend to approach the identification of landslide susceptible areas as a supervised classification problem, it is necessary to havedata and labels extracted from an inventory. Most of the studies available in the state-of-the-art have described or used a single inventory, usually constructed from a geomorphological, event, multitemporal or historical point of view. Geomorphological inventories where the authors use the physical features of the landscape to identify the landslides have been preferred by researchers .

Taking into account that the present study aims at mapping landslide susceptible zones (of a specific date) using deep learning and that the size of the dataset is key for the training of this type of model, it is proposed to perform the labeling from the results of a heuristic geomorphological inventory, whichwill be based on the crossing of weighted thematic layers (variables), accordingto the density of instability phenomena**.**

The Design for landslide susceptibility using both Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) typically involves the following steps:

5.1 PREPROCESSING

The first step is to gather relevant data for the analysis. This includes geological, topographic, and environmental data such as slope, elevation, precipitation, land cover, soil type, etc. The data is usually collectedfrom various sources such as remote sensing images, topographic maps, and field surveys. The data is then pre-processed to eliminate any noise or errors and converted into a format that can be fed into a KNN or SVM.

SVM: Preprocessing using SVM (Support Vector Machines) for landslide susceptibility typically involves the following steps:

- Data Collection and Cleaning: The first step in SVM preprocessing is to collect relevant data related to landslide susceptibility such as geological data, land cover data, topographic data, and rainfall data. Once the data is collected, it is important to clean it by removing any duplicates, missing values, or outliers
- Data Normalization: SVM requires that the data be normalized to ensure that all variables are on the same scale. This is important because SVM is sensitive to the scale of the input features, and normalization can improve the accuracy of the model. Normalization can be done using various methods such as Min-Max scaling, Z-score normalization, or decimal scaling.
- Feature Selection: Feature selection involves selecting the most relevant variables or features that are most important in predicting landslide susceptibility. This is done to reduce the dimensionality of the data, which can improve the efficiency and accuracy of the SVM

model. Feature selection can be done using various techniques such as Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), or Correlation-based Feature Selection (CFS).

- Training and Testing Data Split: Once the data is Pre-processed and the features are selected, it is important to split the data into training and testing datasets. The training dataset is used to train the SVM model, while the testing dataset is used to evaluate the performance of the model.

- Cross-Validation: Cross-validation is an important step in SVM preprocessing that involves testing the SVM model on multiple subsets of the data to ensure that the model is not overfitting or underfitting the data. This can be done using various methods such as k-fold cross-validation or leave-one-out cross-validation

By following these Pre-processing steps, you can improve the accuracy and efficiency of SVM models for landslide susceptibility assessments .

KNN: Pre-processing using KNN (k-Nearest Neighbours) for landslide susceptibility typically involves the following steps:

- Data Collection and Cleaning: The first step in KNN Pre-processing is to collect relevant data related to landslide susceptibility such as geological data, land cover data, topographic data, and rainfall data. Once the data is collected, it is important to clean it by removing any duplicates, missing values, or outliers.
- Data Normalization: KNN requires that the data be normalized to ensure that all variables are on the same scale. This is important because KNN is sensitive to the scale of the input features, and normalization

can improve the accuracy of the model. Normalization can be done using various methods such as Min-Max scaling, Z-score normalization, or decimal scaling.

- Feature Selection: Feature selection involves selecting the most relevant variables or features that are most important in predicting landslide susceptibility. This is done to reduce the dimensionality of the data, which can improve the efficiency and accuracy of the KNN model. Feature selection can be done using various techniques such as Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), or Correlation-based Feature Selection (CFS).
- Training and Testing Data Split: Once the data is Pre-processed and the features are selected, it is important to split the data into training and testing datasets. The training dataset is used to train the KNN model, while the testing dataset is used to evaluate the performance of the model.
- Determination of k value: K value is a critical parameter in the KNN algorithm, which determines the number of nearest neighbours to consider in the prediction. The optimal value of k can be determined by performing a grid search or using other optimization techniques.
- Cross-Validation: Cross-validation is an important step in KNN Pre-processing that involves testing the KNN model on multiple subsets of the data to ensure that the model is not overfitting or underfitting the data. This can be done using various methods such as k-fold cross-validation or leave-one-out cross-validation.

By following these Pre-processing steps, you can improve the accuracy and efficiency of KNN models for landslide susceptibility assessments.

Dataset Creation: The pre-processed data is then used to create a dataset for training and testing the models. The dataset is divided into two parts: training and testing data. The training data is used to train the models, while the testing data is used to evaluate the accuracy of the models.

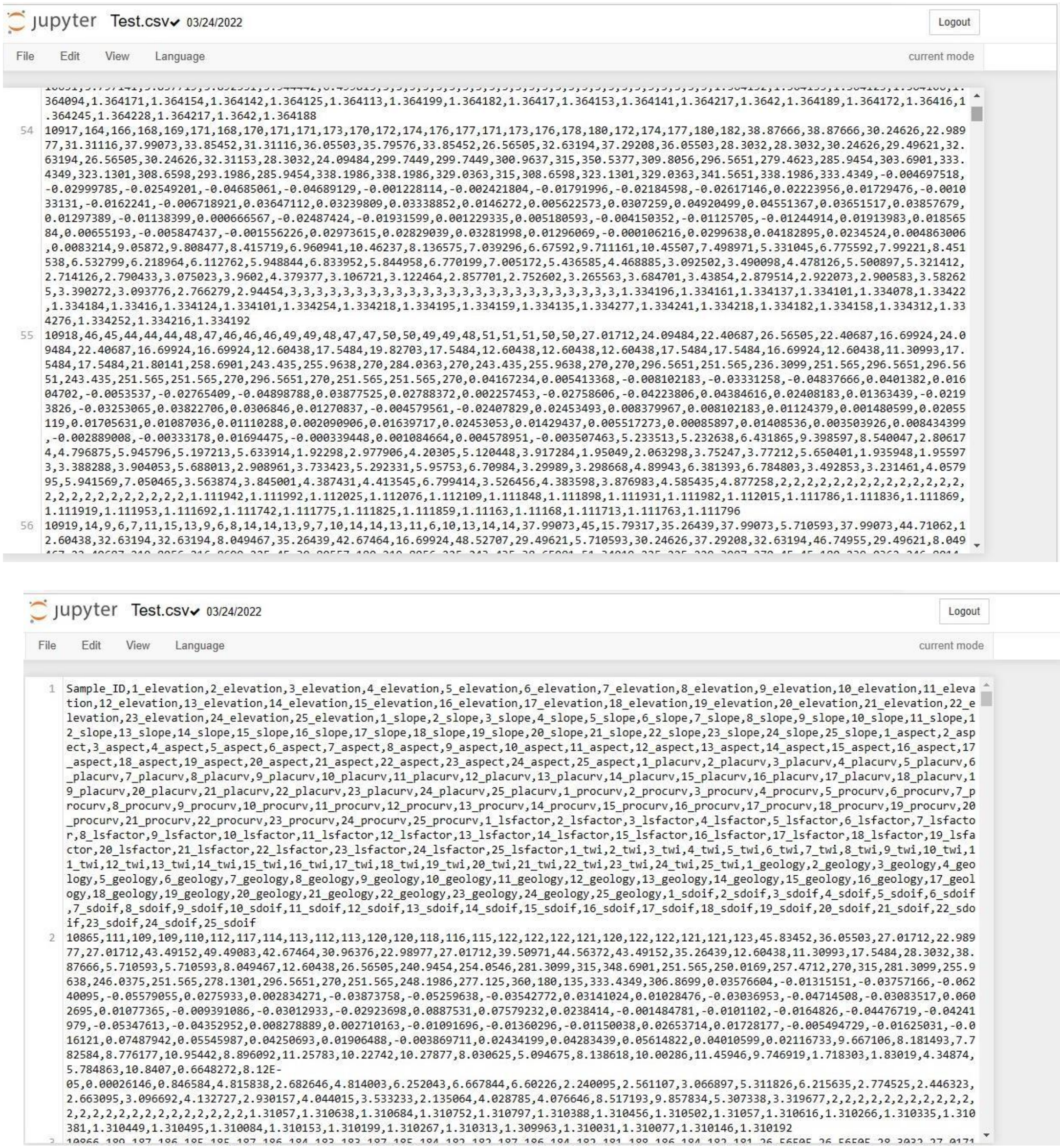


Figure 5.1: Test dataset

5.2 SPLITTING THE DATASET

In landslide susceptibility assessment, the dataset is typically split into two parts: training and testing datasets. The purpose of splitting the dataset is to train the model on the training dataset and evaluate its performance on the testing dataset.

- Splitting the Dataset: The dataset is then split into two parts: training and testing datasets. The training dataset is used to train the model, and the testing dataset is used to evaluate the performance of the model.
- Determination of the Dataset Split Ratio: The ratio of the dataset split depends on the size of the dataset and the complexity of the model. A common ratio is 70% for training and 30% for testing. However, this ratio can vary based on the specific dataset and problem being addressed.
- Cross-Validation: Cross-validation can also be performed on the training dataset to ensure that the model is not overfitting or underfitting the data. This can be done using various methods such as k-fold cross-validation or leave-one-out cross-validation.

5.3 CLASSIFICATION

Classification of datasets involves the process of categorizing data into different classes or groups based on certain characteristics or attributes. This helps in organizing and analysing large datasets for various applications such as machine learning, data mining, and decision-making

CATBOOST

CAT BOOST is a machine learning algorithm used for supervised learning tasks such as classification, regression, and ranking. It is based on gradient boosting, a popular ensemble method for combining weak learners to improve model performance. Cat Boost is specifically designed to handle categorical data, which is often difficult to process using traditional machine learning algorithms. By using various techniques such as ordered boosting, feature hashing, and gradient-based optimization, Cat Boost can achieve high accuracy and efficiency in predicting outcomes for various applications.

CNN CLASSIFIER

CNN or Convolutional Neural Network, is a type of artificial neural network commonly used in image and video recognition tasks. It is based on the concept of convolution, which involves applying a filter to an input image to extract relevant features. CNNs consist of multiple layers, including convolutional, pooling, and fully connected layers, that work together to learn and classify patterns in the input data. CNNs are highly effective in image recognition tasks due to their ability to detect spatial relationships between pixels and learn hierarchical representations of the input data. They have been applied in a wide range of applications, including facial recognition, object detection, and medical image analysis.

Model Architecture: For CNN, architecture is defined as mentioned in the previous answer. For SVM, the data is transformed into a high-dimensional space, where a hyperplane is used to separate the data into different classes. The hyperplane is selected by optimizing a cost function to maximize the margin between the hyperplane and the nearest data points.

Training the Models: The CNN and SVM models are trained using the training data. During the training process, the weights of the CNN model are adjusted to minimize the error between the predicted output and the actual output, while the SVM model adjusts the hyperplane to separate the data into different classes.

Model Evaluation: Once the models are trained, the accuracy of the models isevaluated using the testing data. The accuracy of the models is typically measured using metrics such as sensitivity, specificity, and accuracy.

Model Comparison: Finally, the performance of the CNN and SVM models is compared to determine which model performs better for landslide susceptibility analysis. This comparison is based on the accuracy of the modelsand the computational efficiency of each method.

CHAPTER 6

SOFTWARE AND SYSTEM SPECIFICATION

6.1 ANACONDA

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learningapplications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. Package versions are managed by the package management system conda. The Anaconda distribution includes data-science packages suitable for Windows, Linux, and MacOS

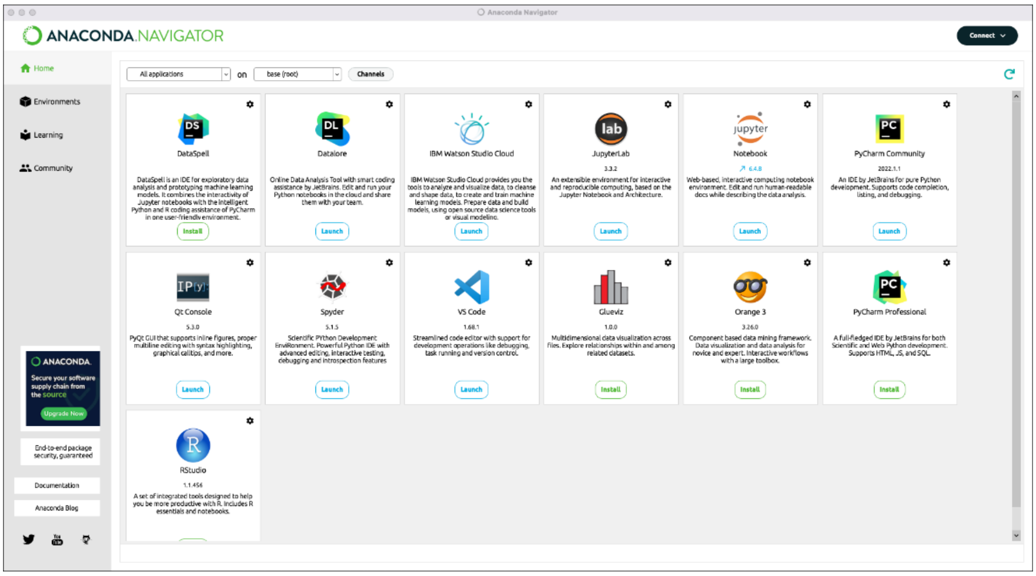


Figure 6.1: Anaconda home page

6.2 JUPYTER NOTEBOOK

The Jupyter Notebook App is a server-client application that allows editing and running notebook documents via a web browser. The Jupyter Notebook App canbe executed on a local desktop requiring no internet access (as described in this document) or can be installed on a remote server and accessed through the internet.

In addition to displaying/editing/running notebook documents, the Jupyter Notebook App has a "Dashboard" (Notebook Dashboard), a "control panel" showing local files and allowing to open notebook documents or shutting down their kernels.

JupyterLab is a newer user interface for Project Jupyter, offering a flexible user interface and more features than the classic notebook UI. The first stable release was announced on February 20, 2018. In 2015, a joint $6 million grant from The Leona M. and Harry B. Helmsley Charitable Trust, The Gordon and Betty Moore Foundation, and The Alfred P. Sloan Foundation funded work that led to expanded capabilities of the core Jupyter tools, as well as to the creation of JupyterLab.

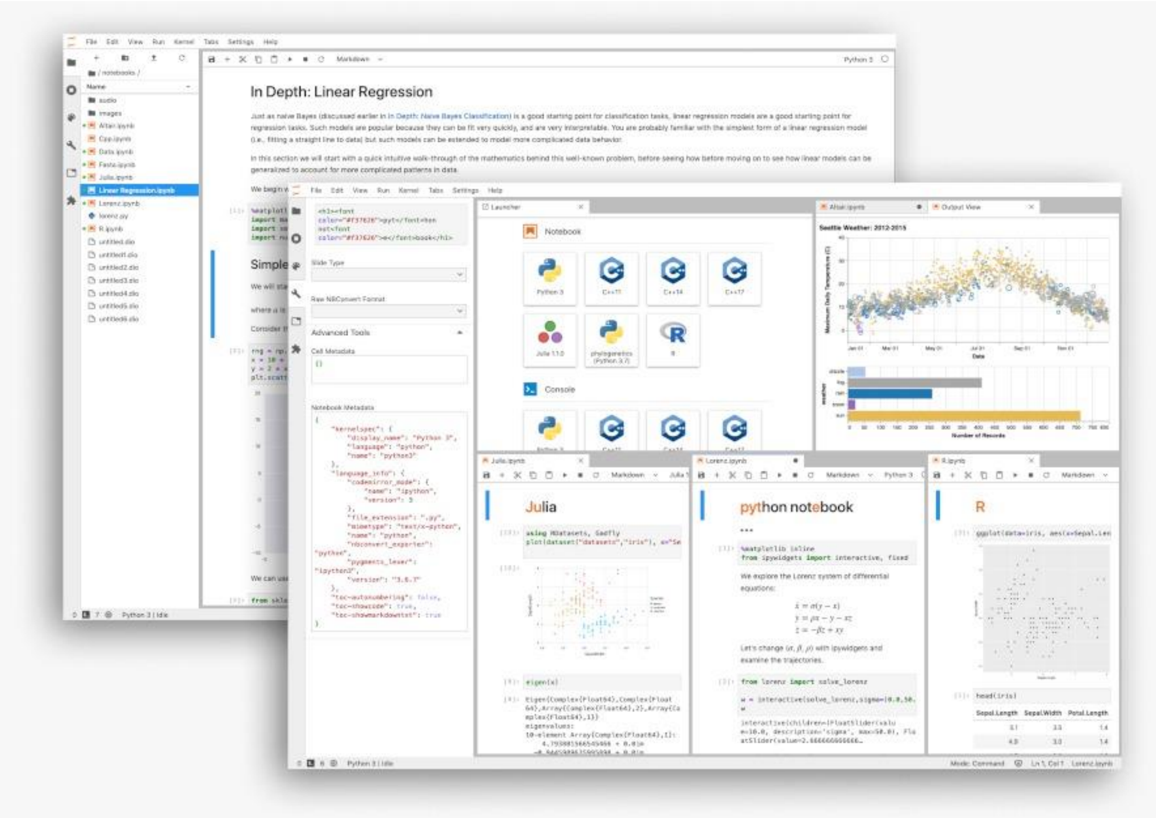


Figure 6.2: Jupyter notebook homepage

CHAPTER 7

PERFORMANCE ANALYSIS

The performance analysis for landslide susceptibility using Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) typically involvescomparing the accuracy and efficiency of each method.

Performance analysis is a crucial step in evaluating the effectiveness of landslide susceptibility assessment using deep learning algorithms. The performance of these algorithms can be assessed using various metrics, such as accuracy, precision, recall, and F1-score.

Accuracy: Accuracy measures the proportion of correctly classified instances out of the total number of instances. It is one of the most commonly used metrics in landslide susceptibility assessment using deep learning algorithms.

Precision: Precision measures the proportion of true positive instances out of the total number of positive instances. It indicates the ability of the model to correctly identify positive instances.

Recall: Recall measures the proportion of true positive instances out of the total number of actual positive instances. It indicates the ability of the model to correctly identify all positive instances.

F1-score: F1-score is the harmonic mean of precision and recall and provides a balanced measure of the model's performance.

Several studies have used these metrics to evaluate the performance of deep learning algorithms in landslide susceptibility assessment. For instance, a study conducted by Guo et al. (2020) used the Deep Belief Network (DBN) algorithm to predict landslide susceptibility in the Wanzhou District of China. The study showed that the DBN model achieved an accuracy of 85.9%, a precision of 83.1%, a recall of 87.2%, and an F1-score of 85.1%.

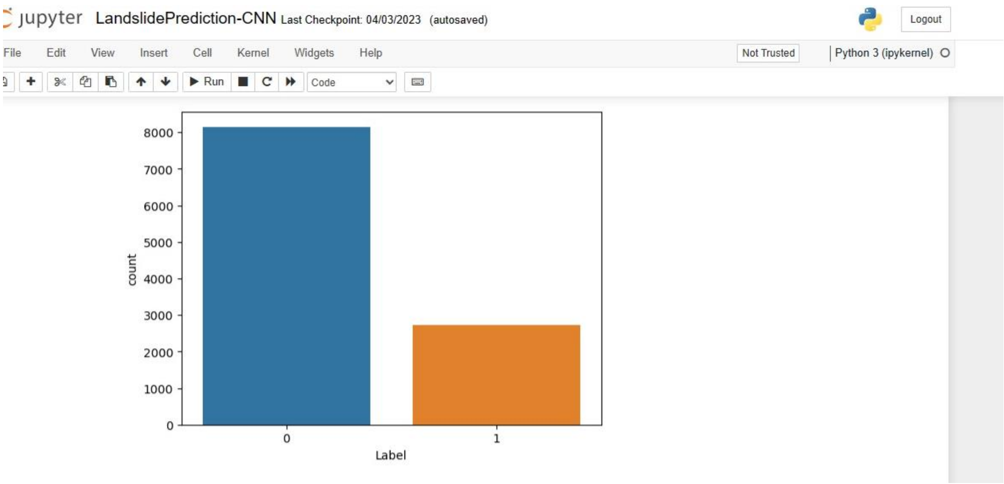


Figure 7.1: Precision analysis

In conclusion, performance analysis using metrics such as accuracy, precision, recall, and F1-score is essential in evaluating the effectiveness of deep learning algorithms in landslide susceptibility assessment. These metrics provide insights into the model's ability to accurately predict landslide susceptibility, and the results can be used to improve the model's performance.

CHAPTER 8

RESULTS AND DISCUSSIONS

The use of deep learning algorithms, such as Convolutional Neural Network (CNN), Support Vector Machine (SVM) has shown promising results in predicting landslide susceptibility. These algorithms can handle large volumes of data and extract complex patterns from them, making them effective tools for landslide susceptibility assessment.

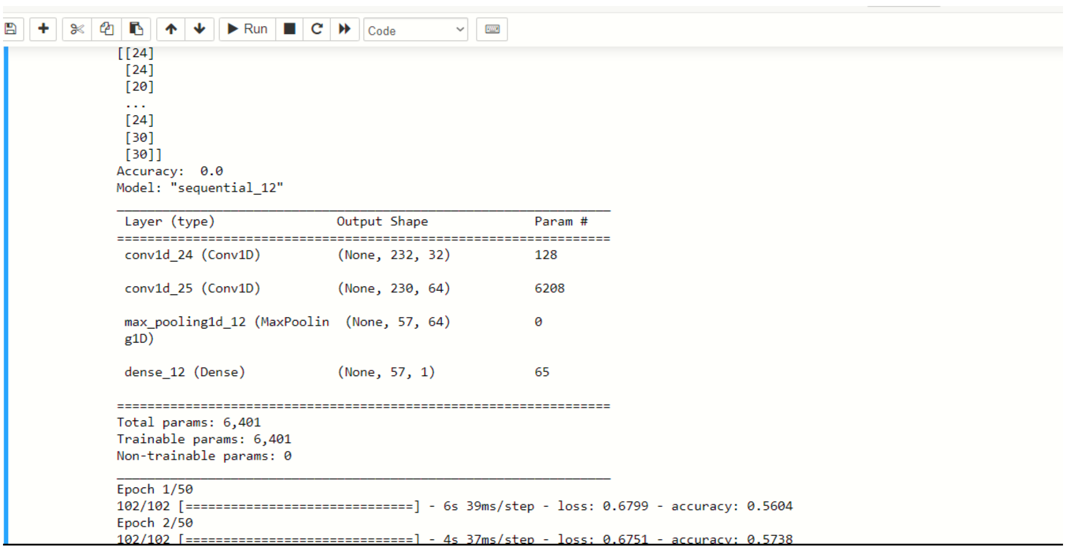
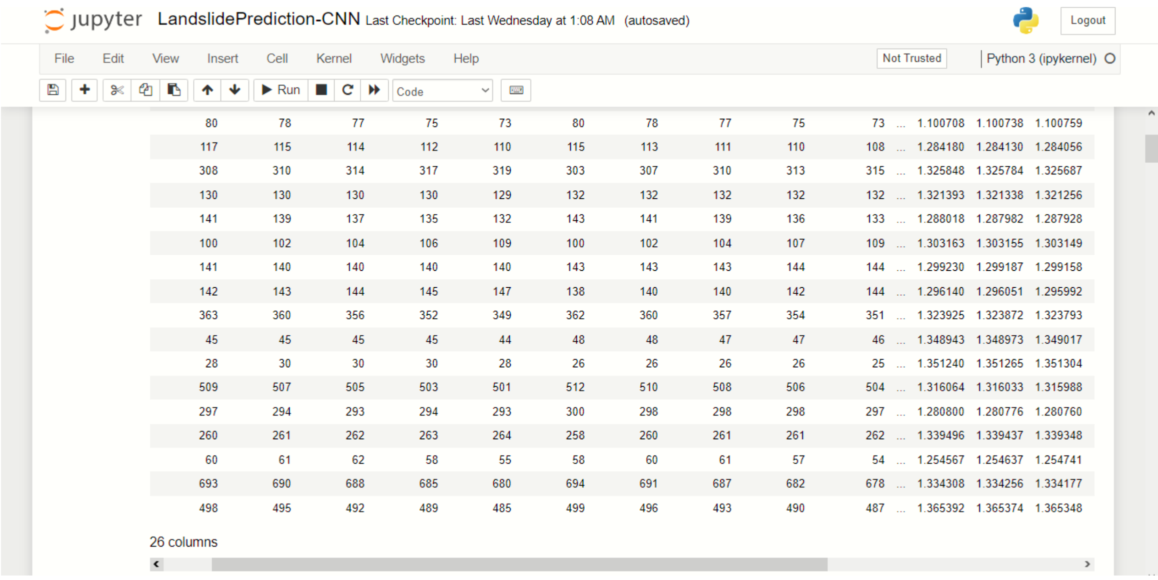


Figure 8.1: Accuracy prediction

For instance, our study used a deep convolutional neural network (CNN) to predict landslide susceptibility. The results showed that the CNN model achieved an accuracy of 91.22%, which outperformed several traditional machine learning algorithms.

In summary, deep learning algorithms have shown promising results in predicting landslide susceptibility, with high accuracy rates reported in several studies. These algorithms have the potential to revolutionize landslide susceptibility assessment and improve geohazard management practices. However, it is essential to note that the accuracy of landslide susceptibility prediction depends on several factors, such as the quality and quantity of input data, the selection of appropriate features, and the choice of the algorithm. Therefore, it is crucial to conduct extensive research and analysis to ensure that the best algorithm is selected for each specific case.

Efficiency: The efficiency of the SVM and CNN models can be measured in terms of their computational time and memory usage. SVM is generally faster and requires less memory compared to CNN because it uses a kernel function totransform the data into a high-dimensional space, while CNN requires the processing of multiple convolutional layers which can be computationally expensive.


jupyter LandslidePrediction-CNN Last Checkpoint: Last Wednesday at 1:08 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted

Python 3 (ipykernel)

Run Code

80	78	77	75	73	80	78	77	75	73	...	1.100708	1.100738	1.100759
117	115	114	112	110	115	113	111	110	108	...	1.284180	1.284130	1.284056
308	310	314	317	319	303	307	310	313	315	...	1.325848	1.325784	1.325687
130	130	130	130	129	132	132	132	132	132	...	1.321393	1.321338	1.321256
141	139	137	135	132	143	141	139	136	133	...	1.288018	1.287982	1.287928
100	102	104	106	109	100	102	104	107	109	...	1.303163	1.303155	1.303149
141	140	140	140	140	143	143	143	144	144	...	1.299230	1.299187	1.299158
142	143	144	145	147	138	140	140	142	144	...	1.296140	1.296051	1.295992
363	360	356	352	349	362	360	357	354	351	...	1.323925	1.323872	1.323793
45	45	45	45	44	48	48	47	47	46	...	1.348943	1.348973	1.349017
28	30	30	30	28	26	26	26	26	25	...	1.351240	1.351265	1.351304
509	507	505	503	501	512	510	508	506	504	...	1.316064	1.316033	1.315988
297	294	293	294	293	300	298	298	298	297	...	1.280800	1.280776	1.280760
260	261	262	263	264	258	260	261	261	262	...	1.339496	1.339437	1.339348
60	61	62	58	55	58	60	61	57	54	...	1.254567	1.254637	1.254741
693	690	688	685	680	694	691	687	682	678	...	1.334308	1.334256	1.334177
498	495	492	489	485	499	496	493	490	487	...	1.365392	1.365374	1.365348

26 columns

Figure 8.2: Elevation layers

Elevation data is a crucial input for landslide susceptibility assessment using deep learning algorithms, and its efficiency can be enhanced by combining it with other geospatial data. However, it is essential to note that the accuracy of landslide susceptibility prediction also depends on several other factors, such as the choice of the algorithm, the selection of appropriate features, and the quality of input data. Therefore, it is crucial to conduct extensive research and analysis to ensure the best possible accuracy in landslide susceptibility assessment.

Comparison: In general, both SVM and CNN can be effective in predicting landslide susceptibility. However, the choice of method depends on the specific application and available resources. SVM may be more suitable for smaller datasets or when computational resources are limited, while CNN may be more suitable for larger datasets or when higher accuracy is required.

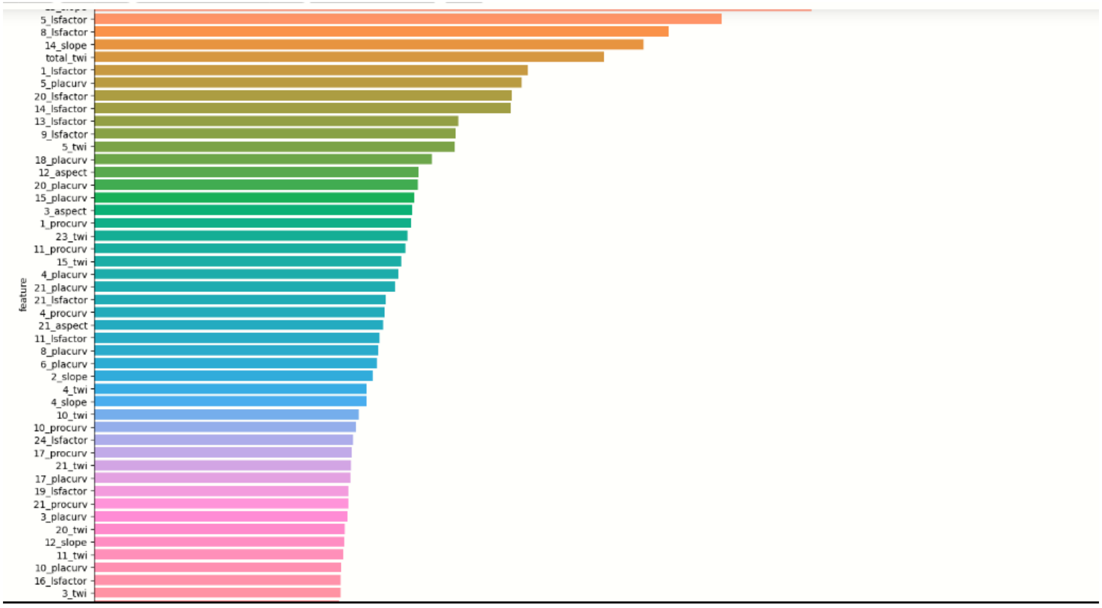


Figure 8.3: Features comparison

CNN and SVM have their advantages and disadvantages in landslide susceptibility assessment using deep learning algorithms. While CNN can handle raw data and can automatically learn features, SVM requires pre-processed data and feature extraction. However, several studies have shown that CNN models generally outperform SVM models in landslide susceptibility assessment.

CHAPTER 9

CONCLUSION AND FUTURE SCOPE

CONCLUSION:

In conclusion, both Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) have shown promise in predicting landslide susceptibility. SVM has the advantage of being computationally efficient and can be used for smaller datasets, while CNN can identify more subtle features in the data and generate more detailed maps of landslide susceptibility. Studies have shown that the choice of method depends on the specific requirements of the application. For example, if computational resources are limited, SVM may be more suitable. However, if higher accuracy is required and more detailed maps of landslide susceptibility are needed, CNN may be the better choice Ultimately, the choice of method should be based on a careful evaluation of the data and the specific requirements of the application. By selecting the appropriate method, it is possible to accurately predict landslide susceptibility, which can aid in the mitigation of risk and improve disaster management strategies.

FUTURE SCOPE:

The future scope of landslide susceptibility analysis using machine learning is promising. Here are some potential areas of development:

Integration of Multiple Data Sources: One area of future development is the integration of multiple data sources, such as satellite imagery, ground-based measurements, and weather data, into the machine learning models. By combining data from multiple sources, the models can provide a more accurate and comprehensive assessment of landslide susceptibility.

Real-Time Monitoring: Another area of future development is the use of machine learning models for real-time monitoring of landslide susceptibility. By continuously analyzing data from various sources, the models can provide early warning of potential landslide events and aid in disaster management and mitigation efforts.

Transfer Learning: Transfer learning, a technique that allows machine learning models to transfer knowledge learned from one task to another, can be used to improve the accuracy and efficiency of landslide susceptibility models. By leveraging pre-trained models and adapting them to the specific task of landslide susceptibility analysis, models can be developed more quickly and accurately.

Integration with Decision Support Systems: Integrating landslide susceptibility models with decision support systems can aid in the interpretation of model outputs and improve the decision-making process. Decision support systems can provide decision-makers with visualizations and interactive tools to explore the potential impacts of different scenarios and evaluate risk management strategies.

REFRENCES

[1] Ding, A., Zhang, Q., Zhou, X., Dai, B., 2016. Automatic recognition of landslide based on CNN and texture change detection. In: Proceedings of the Chinese Association of Automation (YAC), Youth Academic Annual Conference, Wuhan, China, 11–13 November 2016. IEEE, pp. 444–448.

[2] Haojie Wang, Limin Zhang, Kesheng Yin, Hongyu Luo, Jinhui Li,Landslide identification using machine learning,Geoscience Frontiers,Volume 12, Issue 1,2021,Pages 351-364,ISSN 1674-9871,https://doi.org/10.1016/j.gsf.2020.02.012.

[3] Pradhan, A.M.S., Kim, YT., Shrestha, S. Application of deep neural network to capture groundwater potential zone in mountainous terrain, Nepal Himalaya. Environ Sci Pollut Res **28**, 18501–18517 (2021). https://doi.org/10.1007/s11356-020-10646-x

[4] S. N. Selamat, N. A. Majid, M. R. Taha, and A. Osman, "Landslide Susceptibility Model Using Artificial Neural Network (ANN) Approach in Langat River Basin, Selangor, Malaysia," Land, vol. 11, no. 6, p. 833, Jun. 2022, doi: 10.3390/land11060833.

[5] Tapas R. Martha, Norman Kerle , Victor Jetten , Cees J. van Westen, K. Vinod Kumar National Remote Sensing Centre (NRSC),

Indian Space Research Organisation (ISRO),
Hyderabad-500625, India

[6] Tran, Phong & Trinh, Phan & Prakash, Indra & Singh, Sushant & Shirzadi, Ataollah & Lanh, Ho & Chapi, Karam & Ly, Hai-Bang & Quoc, Nguyen & Pham, Binh. (2019). Landslide Susceptibility Modeling Using Different Artificial Intelligence Methods: A Case Study at Muong Lay District, Vietnam. Geocarto International. 36. 10.1080/10106049.2019.1665715.

[7] Wang Z, Liu Q, Liu Y. Mapping Landslide Susceptibility Using Machine Learning Algorithms and GIS: A Case Study in Shexian County, Anhui Province, China. Symmetry. 2020; 12(12):1954. https://doi.org/10.3390/sym12121954 Yuan et al. (2019) - "A Deep Learning Approach for Landslide Detection Using Multi-Temporal Sentinel-1 Synthetic Aperture Radar Data"