International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

Inquisition of Landslide Hazards in Tirupattur District, South India Using Geospatial and Machine Learning Techniques

Sindhu G, Rathi S†

PG Student, Department of Computer Science and Engineering, Government College of Technology, Coimbatore- 641013, Tamil Nadu, India.
Professor, Department of Computer Science and Engineering, Government College of Technology, Coimbatore- 641013, Tamil Nadu, India.

Submitted: 15-07-2022 Revised: 25-07-2022 Accepted: 27-07-2022

ABSTRACT
The present study aims at Inquisition of landslide hazards in Tirupattur district, south india using geospatial and machine learning techniques. The major objective specifies in evaluating the capabilities of Four advanced machine learning techniques (MLTs), including, Support Vector Machine (SVM), Random Forest (RF), Ensemble – Boosted trees, Artificial Neural Network (ANN) which incorporates Geographical Info System (GIS) and Python open-source software. Landslides are resonances of rock and soil and the impact of it causes immense human and economic losses. 24 randomly segregated landslide areas were separated into two groups with a ratio of 70% for training and 30% for validating purpose. Sixteen parameters were considered for landslide susceptibility modeling, which includes Altitude, Lithology, Distance to Faults, Normalized Difference Vegetation Index (NDVI), Land Use/Landcover (LULC), Distance to Roads, Slope Angle, Distance to Streams, Soil, Geology, Geo Morphology, Lineament, Drainage, Plan Curvature, Slope Length (LS) and Slope-Aspect. The area under curve (AUC-ROC) has been applied to evaluate, validate and compare the MLTs performance by 25 different aspects. The results 25 indicated that AUC values for two MLTs range from 96.5% for Ensemble – Boosted trees and 95.3% form Support Vector Machine (SVM). The reflection of this study and the landslide 25 susceptibility maps would be useful for environmental conservancy.

Keywords: Landslide susceptibility, Machine learning algorithms, Variable's importance, Geospatial Techniques, South India.

I. INTRODUCTION

Landslides are the common mass wasting processes along the mountainous regions of Tirupattur District, South India, which connects major cities through different modes of transportation (Youssef et al., 2012; Youssef and Maerz, 2013; Youssef et al., 2013, 2014a, b; Elkadiri et al., 2014; Maerz et al., 2014). The nature of landslides such as rock falls, rock and soil sliding and debris flows in India are usually triggered by intense rainfall (Youssef et al., 2013, 2014c). Since the areas are tend to be steep slope and ghat regions, it has several challenges to carry out proper urbanization in the future days. It is true in India, where thousands of people live in this mountainous region and commute along escarpment highways, which consider high-risk threat areas related to landslides. Therefore, it is essential for such a type of area to assess and prevent landslide disasters.

Landslides represents natural hazards in and around the mountainous regions causing mortality, property damage, and consequently economy crisis. These landslides are triggered due to various external factors like heavy rainfall, earthquakes, volcanoes and anthropogenic activities (Guzzetti et al., 2007; Lin et al., 2007; Hadi et al., 2018; Roback et al., 2018; Strupler et al., 2018; Gordo et al., 2019; Roccati et al., 2019). In order to violate these problems, the landslide susceptibility maps play a crucial role in determining the most vulnerable areas for landslides. This paradigm can be articulated using different landslides-conditioning factors (e.g.,

lithology, lineaments, geomorphology, soil type and depth, slope angle, slope aspect, curvature, altitude, engineering properties of the lithological material, land use patterns, and drainage networks). Various studies have been carried out on landslide susceptibility assessment using Remote Sensing and GIS techniques (e.g., Saha et al., 2005; Pradhan and Youssef, 2010; Pradhan et al., 2010; Bednarik et al., 2012; Mohammady et Gordo et al., 2019; Nohani et al., 2019; Pham et al., 2019; Pourghasemi et al., 2019a; Roccati et al., 2019; Yan et al., 2019).

Various investigators identified the landslide susceptibility by changes or impact of one layer over another. (Guzzetti et al., 2006; van Westen, 2006; Constantin et al., 2011). Using various paradigm in landslide forecast, the after effects could be decreased to a certain extent (Pradhan, 2010). Landslide occurrence will be predicted in prior and controlled with the use of landslide inventories with the required data. In the prognostic approaches based on earlier landslide events, the basic premise is that future landslides will have similar distribution patterns to those that occurred in the past (i.e., the past is the key to the future; reversed Uniformitarianism). The future landslide prediction can be produced by the susceptibility models with the available statistical data. Susceptibility models represent the prospect of a landslide occurring in any specific location in terms of relative probability (high, medium, low). Susceptibility maps can be obtained for different types of natural hazards through several methods: (a) direct mapping of susceptibility spot based on expert criteria (Ardau et al., 2007; Razak et al., 2011; Salvatici et al., 2108). These maps are subjective and depend on the expert perceptions; (b) the deterministic models based on the stability analysis (mathematical relationships between resisting and driving forces) predicts the unsafe regions considering few constraints.(e.g., engineering characteristics of the rocks and soils, slope geometry, discontinuity characteristics, and hydrological conditions) (Yilmaz, 2009). (b) the deterministic models are based on stability analyses (mathematical relationships between resisting and driving forces) that take into consideration a number of parameters involved in the hazard processes (e.g., engineering characteristics of the rocks and soils, slope geometry, discontinuity characteristics, and hydrological conditions) (Yilmaz, 2009). These methods can be used efficiently to a single-slope scale (Ayalew and Yamagishi, 2005). However, the main consequences of these models are that streamlining the complexity of the hazard processes, generally static, incorporate geometrical suppositions, and difficult and expensive to acquire geotechnical and hydrological data especially when examining large and heterogeneous areas; (c) the heuristic approach involves establishing susceptibility classes by judging the relative contribution of a number of variables on landslide formation (e.g. Dai and Lee, 2002; Dahal et al., 2008a, 2008b). The main shortcoming of the heuristic methods is the subjective component of the susceptibility assessments (Dai et al., 2001); (d) the probabilistic methods allow the of susceptibility models erection which analyses the statistical relationships between the spatial distribution of known landslides and that of a set of controlling factors. Several methods have been carried out to assess landslide susceptibility assisting GIS techniques. Few contemporary researches lately give the data of bivariate analyses, quantifying the spatial relationships between landslide and specific variables governing their distribution (e.g., Constantin et al., 2011; Jaafari et al., 2014; Pradhan et al., 2014; Regmi et al., 2014; Razandi et al., 2015; Zhang et al., 2016; Ding et al., 2017; Shirzadi et al., 2017; Sestrașet al., 2019). Further researches in the present days looks upon the multivariate logistic regression methods (Park et al., 2013; Shahabi et al., 2015; Tien Bui et al., 2016); knowledge-based methods (Althuwaynee et al., 2016; Kumar and Anbalagan, 2016), and multivariate binary logistic regression (Mandal et al., 2018); (e) Various machine learning models were employed in landslide susceptibility mapping such as support vector machine - SVM (e.g., Tien Bui et al., 2012; Pourghasemi et al., 2013c; Hong et al., 2015; Colkesen et al., 2016; Lee et al., 2017; Kalantar et al., 2018; Pourghasemi and Rahmati, 2018), fuzzy logic-FL (Kumar and Anbalagan, 2015; Shahabi et al., 2015), artificial neuronal networks-ANNs (e.g., Pradhan and Lee, 2010; Yilmaz, 2010; Zare et al., 2013; Elkadiri et al., 2014; Pham et al., 2015; Arnone et al., 2016; Gorsevski et al., 2016; Wang et al., 2016; Aditian et al., 2018; Pourghasemi and Rahmati, 2018), neuro-fuzzy-NF and adaptive neuro-fuzzy inference system-ANFIS (e.g., Sezer et al., 2011; Dehnavi et al., 2015; Aghdam et al., 2016; Nasiri Aghdam et al., 2016; Chen et al., 2019), decision tree-DT (e.g., Kavzoglu et al., 2014a,b; Wu et al., 2014; Tien Bui et al., 2016), generalized additive model-GAM (e.g., Vorpah et al., 2012; Goetz et al., 2015; Pourghasemi and Rahmati, 2018), adaBoost-AB (Micheletti et al., 2014), random forest-RF (e.g., Paudel and Oguchi, 2014; Trigila et al., 2015; Hong et al., 2016a; Pourghasemi and Kerle, 2016; Youssef et al., 2016; Chen et al., 2017; Taalab et al., 2018; Park and Kim, 2019), naïve bayes-NB (e.g., Pham et al., 2015;

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

Tsangaratos and Ilia, 2016; Pham et al., 2017; Chen et al., 2018a,b), kernel logistic regression-KLR (Tien Bui et al., 2016; Chen et al., 2018b), boosted regression tree-BRT (Dickson and Perry, 2016; Youssef et al., 2016; Park and Kim, 2019), classification and regression tree-CART (Vorpah et al., 2012; Felicísimo et al., 2013; Youssef et al., 2016; Chen et al., 2017; Pourghasemi and Rahmati, 2018), general linear model-GLM (e.g., Youssef et al., 2016; Pourghasemi and Rahmati, 2018), multivariate adaptive regression splines-MARS (Vorpah et al., 2012; Felicísimo et al., 2013; Conoscenti et al., 2015), maximum entropy-MaxEnt (Felicísimo et al., 2013; Park, 2015; Hong et al., 2016b; Kornejady et al., 2017), and quadratic discriminant analysis-QDA (Rossi et al., 2010; Pourghasemi and Rahmati, 2018).

II. STUDY AREA

As per the data from The Survey of India (SOI) Toposheet (SOI, 2011) Number is 57 L/8, 57 L/9, 57 L/10 and 57 L/11 with latitude $12^0 35'0"$ - $12^0 55'0"$ N and longitude $78^0 30'00"$- $78^0 50'00"$ E (Fig. 1) which is located in Northern part of Tamil Nadu, South India. The overall geographical area is 1727 Km^2 and elevation ranges from 220 m to 351 m above the Mean Sea Level (MSL) (Venkatesan and Subramani 2019). This terrain has a semi-arid climate, 135 km west of the state capital Chennai. The topography is almost plain with slopes from west to east and Palar river basin streams from Andhra Pradesh and the streamflow enters Tirupattur district at Kanaganachiamman Koil, Nattrampalli Taluk which lies in the Eastern Ghats region of Tamil Nadu. The temperature ranges from 13 °C (55 °F) to 39.4 °C (102.9 °F). Like the rest of the state, April to June are the hottest months and December to January are the coldest. The most common types of hard rock formations in this area are the Gneisses and Charnockites. The Gneissic formations are found in almost all the taluks of the district but lack uniformly both in composition and texture. Different names are attributed to the Gneissic formation based on its mineral content.The sedimentary or the Quarternary Alluvial deposits which are the transported sediments by the river and streams stretch mainly along the Palar river course as thin isolated patches. These formations overlie the Hard rock formation. (Venkatesan et al. 2020c).

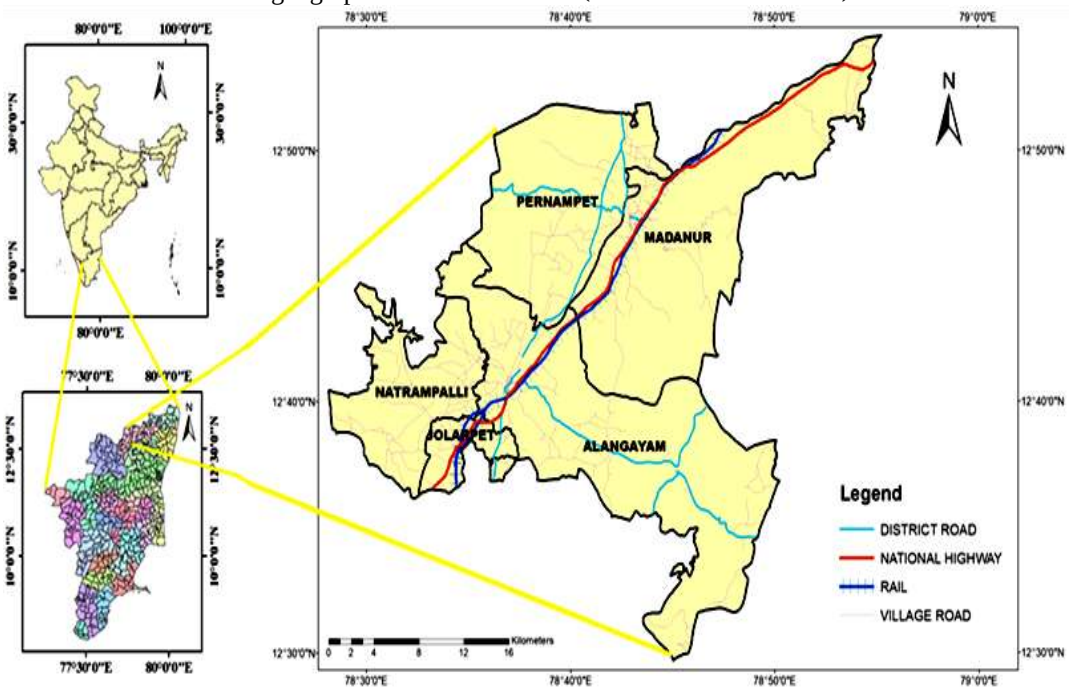


Fig. 1 Study area map

III. MATERIALS AND METHODS

The current approach demonstrates the landslide susceptibility mapping using different data sources. The detailed methodology of the study is presented in the Fig. 2.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

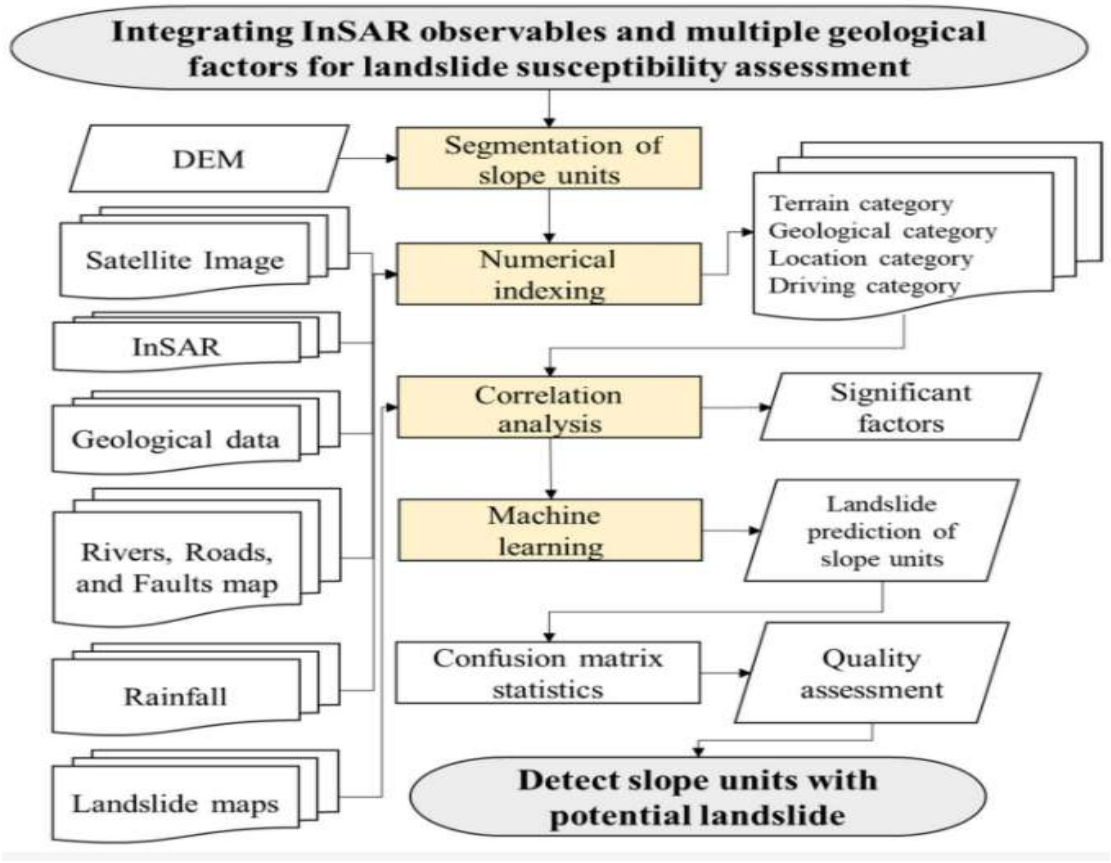


Fig. 2 Detailed Methodology

The landslide prediction zones in the study area were demarcated by integrating different thematic maps using remote sensing and GIS techniques. Geological map of the area was prepared from the Geological Quadrangle Map published by the Geological Survey of India (GSI). Other thematic maps such as geomorphology, drainage, lineament, land use / land cover, soil and slope were prepared from the satellite data (LANDSAT image of Path: 102 and Row: 125, 2018, Earth Explorer and SRTM) using GIS. By integrating all the thematic layers landslide model was generated. Integration of various thematic maps was carried out using GIS in the following three steps:

Spatial database building

By digitizing scanned maps, editing for errors, topology building, attribute assignment, and projection, all the thematic maps were generated using GIS software. (Sharma et al. 2012).

Spatial data analysis

The landslide predict index map was prepared for the study area by incorporating the hydrogeomorphic, linear, geological, slope, and soil maps along with drainage patterns (Mohanty and Behrera 2010). Based on its degree of impact on land use and land cover each theme was considered and assigned a weightage. These predictions were prepared by combining all these thematic maps and adding restricted slope level data. Different geological formations that develop a range of landforms such as structural hills, pediments, buried pediments, and valley fills have different water keeping abilities that show varied qualities of landslide

Data integration

Each thematic map provides some information on the occurrence of landslide, such as geology, geomorphology, drainage, drainage density, line, line density, soil, slope, and land use/land cover. To find the intersecting polygons, each theme was overlaid on another theme. A new map was obtained by combining two thematic maps in this way. This composite map was subsequently overlaid, and so on, on a third thematic map. So there was a final composite map made. Weighting was allocated to each polygon in the final map using a basic arithmetic model. (Schot & Vander wal 1992, Mohanty & Behrera 2010).

Numerical Indexing and Corelation Analysis

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

Terrain, Geological, Location and Driving categories were analysed with the data available from various segregations like In - SAR Satellite Image, Geological Data, Rainfall Data and River, Roads and Fault Maps. The outcome of Numerical Indexing will be corelated with the aid of landslide maps and the analysis will be done considering significant factors.

Machine Learning, Matrix Statistics & Quality Assessment

The landslide prediction of slope units will be obtained by applying Machine Learning Techniques like SVM, Boosted Tree approaches. The Confusion Matrix Statistics are obtained from the prediction of different slope units. The datasets are then assessed and final prediction is evaluated to detect slope units with potential landslide.

3.1. Data collection

Various data sources include field data, historical data acquired from civil defense authority, 202 national reports, and questionnaires with dwellers were collected to identify the old landslide events and their frequencies. Satellite images including enhanced thematic mapper plus (ETM+ 203) 204 (acquired in April 2016, spatial-resolution of 15 m (after image fusion of 30 m image composite 205 with panchromatic band 15 m)), high-resolution images (GeoEye images with a spatial206 resolution of 0.5 m which acquired in May 2016, and professional Google Earth data with a 207 resolution less than 1 meter), and DEM derived from 1:10,000 topographic contours using 208 triangulated irregular network (TIN) data structures (spatial resolution 10 m) were used in this 209 study. Geological map (Abha quadrangle, GM-75) with a scale of 1: 250,000 was scanned and 210 converted to a digital format. Finally, these data sets were projected according to UTM-Zone 38, 211 WGS84 Datum.

3.2. Landslide inventory map

To prepare landslide susceptibility mapping, two datasets are essentially required. The first 214 dataset represents the landslide inventory map. The second dataset is related to landslide 215 conditioning factors. Landslide inventory data need to be evaluated against landslide 216 conditioning factors to determine their significance and effect on the occurrence of the landslides 217 because it is typically assumed that landslides will occur under the same conditions as before. 218 Performing a landslide susceptibility model needs a clear understanding of the relationship 219 between the real landslides (landslide inventory dataset) and the landslide conditioning factors 220 (Ercanoglu and Gokceoglu, 2004; van Westen et al., 2006; Petley, 2008). In the current study, an 221 inventory map of existing landslides was prepared by integration of the historical records, field 222 investigations, and data extracted from satellite images interpretation. Historical landslides can 223 be detected from high satellite imageries according to their geomorphological features (e.g., breaks in the vegetated area and bare soil, presence of flow materials along gullies, rims, and 225 streams, different erosional features, flow tracks, depositional fans, circular and planar failures) 226 (De La Ville et al., 2002; Youssef et al., 2009, 2016; Elkadiri et al., 2014). Field surveys were 227 carried out to verify the historical and landslide extracted from high resolution images and to 228 collect fresh/new landslides in the study area. These landslides have volumes ranging from few 229 cubic meters to about few hundreds cubic meters. All inventory data were assembled to establish 230 a landslide inventory map. In the current study, a total of 243 landslides were used which 231 distributed all over the study area (Fig. 1). Then, the inventory data must be divided into the 232 training and validating datasets to be used for building the models and verifying processes, 233 respectively (Ohlmacher and Davis, 2003; Chacon et al., 2006). In landslide modeling, there is 234 no specific pre-defined method that exists for categorizing inventory data. It is typically decided 235 based on the accessibility and quality of data. According to the literature, the percentages 236 commonly applied to divide the inventory dataset are 70% and 30% for the training and 237 validating datasets, respectively (Nefeslioglu et al., 2008a; Pourghasemi and Rahmati, 2018). In 238 this study, the landslide datasets were randomly divided into two groups, model training group 239 was prepared using 70% of the inventory sites (173 sites) whereas, model validating group 240 consists of 30% of the inventory sites (70 sites) (Fig. 1). The validation process was executed by 241 comparing the existing landslide locations (validating group) with the acquired landslide 242 susceptibility map.

3.3. Landslide conditioning factors (LCFs)

To achieve high accuracy of landslide susceptibility model in predicting landslide vulnerable 245 areas, selecting and preparing the LCFs database is vitally important step. The LCFs in the 246 current study were selected based on the information collected from the literatures, related to the 12 247 study area, and field investigation (Pourghasemi et al., 2013a, b, c; Tseng et al.,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

2015; Youssef, 248 2015; Keesstra et al., 2016; Tien Bui et al., 2016; Kornejady et al., 2017; Ghorbanzadeh et al., 249 2019; He et al., 2019; Sevgen et al., 2019; Xiao et al., 2019). Goetz et al. (2015) indicated that 250 precipitations and earthquakes are external and triggering factors of landslides. However, the 251 previous data on these triggering factors in relation to landslide occurrence are not available in 252 the current area; therefore, these two triggering factors have not been involved in this study. 253 Tseng et al. (2015) mentioned that the selection of landslide related factors (known as internal 254 factors) depends on the characteristics of the study area, the landslide type, and the scale of the 255 analysis. In the current study, twelve LCFs were chosen including, slope-angle, slope aspect, 256 slope length (LS), distance from roads, lithology, distance from wadis, distance from faults, 257 altitude, normalized difference vegetation index (NDVI), plan curvature, profile curvature, and 258 landuse/landcover. These factors will be used in predicting the landslides prone areas in the area 259 under consideration. A database including these 12 LCFs were generated using geographical 260 information system (GIS) for data interpretation and analysis. Thematic layers with a 20 m × 20 261 m spatial resolution pixel size were prepared (Fig. 4). All layers have UTM coordinate system 262 zone 38 with a Datum of WGS 84. In this work, the LCFs have different types; nominal 263 (categorical) where data do not have a natural order or ranking such as (lithology, slope aspect, 264 and landuse/landcover), and ordinal (continuous data) in which the order matters such as 265 (altitude, slope angle, slope length, NDVI, plan-curvature, profile-curvature, distance from 266 faults, distance from roads, and distance from wadis).

3.3.1. Slope angle

The steeper the slope angle of the slope, the larger the number of the landslides. The stability of 269 the slope against failure can be defined by the relationship between shear forces and the 13 270 resistance to shear (safety factor), where the driving force of mass movement increases with 271 increasing the slope angle (Guillard and Zezere, 2012). Tien Bui et al. (2017) concluded that 272 slope steepness is related both to the shear stresses acting on the hill slope and to the 273 displacement of the landslide mass. Magliulo et al. (2008) indicated that slope gradient play a 274 vital role in subsurface flow and impacts on concentration of the soil moisture which are directly 275 related to the landslide occurrence. The slope angle map of the current study was derived from the Digital Elevation Model (DEM) with a resolution of 20 m, ranging from 0◦ to 59.8◦ 276 (Fig. 4a).

3.3.2. Slope aspect 278 Slope-aspect can impact various processes which have direct and indirect influences on 279 landslides including, wind directions, precipitation patterns, sunlight influence, discontinuities 280 orientations, hydrological processes, evapotranspiration, concentration of the soil moisture, 281 vegetation, and root development (Neuhäuser et al., 2012; Quan and Lee, 2012; Devkota et al., 282 2013). Various studies indicated that there is a correlation between slope aspect and other geo283 environmental factors (landslide-related factors) (Van Den Eeckhaut et al., 2009; Regmi et al., 284 2010). Capitani et al. (2013) applied conditional analysis to all of the possible combinations 285 between the slope aspect and landslide-predisposing factors (lithology, slope angle, distance to 286 the hydrographic elements and distance to the tectonic lineaments). They concluded that the 287 slope aspect significantly influenced the distribution of superficial landslide types, but apparently 288 not that of other landslide types. Slope aspect in the study area was constructed using DEM and 289 classified into nine categories including flat (–1), North (0°–22.5°; 337.5°–360°), North-East 290 (22.5°–67.5°), East (67.5°–112.5°), South-East (112.5°–157.5°), South (157.5°–202.5°), South291 West (202.5°–247.5°), West (247.5°–292.5°), and North-West (292.5°–337.5°) (Fig. 4b).

3.3.3. Slope length (LS)

The LS is another topographic factor that has a crucial impact on landslide susceptibility. Slope 294 length incorporated with slope-angle and affect the soil loss and hydrological processes of the 295 mountain areas (Pourghasemi and Rahmati, 2018). In the current study, the LS factor, ranges 296 from 0 to 35.27 as shown in Fig. 4c, was extracted from DEM using SAGA software according 297 to Eq. (1) (Moore and Burch, 1986):

$$LS = \left(\frac{As}{22.13}\right)^{0.4}\left(\frac{Sin\beta}{0.0896}\right)^{1.3} \quad (1)$$

(1) where, As (m2 299) is specific catchment area and β is in degree.

3.3.4. Distance from roads 302 The establishing of mountain roads (escarpment roads) require engineering activities such as 303 cutting or excavating slopes, leading to changing the original geological conditions and 304 weakening the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

natural support of rock slopes, these activities will have a significant and adverse 305 impact on the landslide occurrences (Wang et al., 2016; Xiao et al., 2019). Road construction 306 (cutting and excavation) is an important anthropogenic factor that influences the topography of 307 natural slopes (Xiao et al., 2019). Accordingly, distance from roads could be potential indicator 308 for the landslides. In the current study, distance from roads has a range (0 – 2,883 m), was 309 developed using the Euclidean distance tool in ArcGIS 10.2 (Fig. 4d). 310

3.3.5. Lithology 311 Lithology variations have a significant impact on different types of geo-hazards (e.g., landslides 312 and land subsidence) (Rahmati et al., 2016). These units vary in physical and mechanical 313 characteristics including, type, strength, degree of weathering, durability, density, and 314 permeability (Henriques et al., 2015). In the current study, the lithology was extracted from the 15 315 geological map (1: 250,000-scale) acquired from the Saudi Geological Survey database. To 316 enhance the resolution of the lithology map, landsat images (15 m resolution) were used to verify 317 the lithology types using enhancement processing techniques (principle component analysis and 318 band ratio combinations). Nine lithological units were identified including, biotite-quartzite319 plagioclase granofels, gabbro, alluvium and gravel, Jeddah and Bahah groups, Bahah group 320 within the Tayyah bet, biotite monzogranite, basalt and andesite, metagabbro and gabbro, and 321 Wajid sandstone (Fig. 4e). 322

3.3.6. Distance from wadis (main streams) 323 In the current study, DEM and topographic map were used to extract wadis (main streams) In 324 fact, runoff along wadis plays a cardinal role in undercutting phenomena and increasing the pore 325 water pressure of the areas adjacent to wadis and initiating landslide (Haigh and Rawat, 2012; 326 Hadji et al., 2013). Therefore, it becomes an essential conditioning factor in landslide 327 susceptibility (Pradhan et al., 2010). Distance from wadis, which has a range (0 – 1,974 m), was 328 extracted using the Euclidean distance operation in ArcGIS 10.2 (Fig. 4f). 329

3.3.7. Distance from faults 330 The presence of structural discontinuities, which are tectonic breaks including faults, folds, 331 fractures, joints, and shear zones, play a vital role in weakening the rock masses (decreasing the 332 rock strength) and causing landslides (Lee et al., 2002; Kanungo et al., 2006; Bucci et al., 2016). 333 So that, distance from faults could be potential indicators for the landslides. In the current study, 334 the geological map 1: 250,000-scale was used to extract the faults. Landsat images (15 m 335 resolution) and high-resolution satellite images (GeoEye 05m and Google earthtechniques of the Landstat images were provided the best visualization of the faults. The 339 Euclidean distance tool in ArcGIS 10.2 was used to produce distance from faults map. Distance 340 from faults has a range of 0 –2,839 m (Fig. 4g). 341

3.3.8. Altitude 342 Altitude is controlled by various geological, geomorphological, and meteorological factors 343 including, lithological units, weathering, wind action, and precipitations (Tsutsui et al., 2007; 344 Vorpahl et al., 2012; Pourghasemi et al., 2013). Feizizadeh et al. (2014) indicated that altitude is 345 one of the topographic factors that influence slope instability. It has been frequently used in 346 almost all landslide susceptibility analysis. In this study, altitude map was prepared according to 347 classification of the built DEM. It ranged between 1,959 m and 2,992 m above sea level (Fig. 348 4h).

3.3.9. Normalized difference vegetation index (NDVI) 350 The NDVI is considered an influencing variable in landslide susceptibility modeling 351 (Althuwaynee et al., 2012). The NDVI plays an important role in immobilizing large amount of 352 water and increasing the shear resistance and soil cohesion of the lithological mass (Sidle and 353 Ochiai, 2006). The NDVI was used in this study to mirror the vegetation density in any area. In 354 general, the value of NDVI ranged from −1 to 1; the high the value the denser the vegetation cover. The NDVI value of the current study area was extracted from the Landsat ETM+ 355 images 356 (acquired in January 2015) based on Eq. (2):

$$NDVI = \frac{(NIR-R)}{(NIR+R)}$$

where NIR and R are the near infrared and red bands of the electromagnetic spectrum. The 359 NDVI values in the study area ranges from (–0.4 to 0.585) where minus portion represents 360 barren land and positive part shows vegetated areas (Fig. 4i).

3.3.10. Plan and profile curvatures 362 Haigh and Rawat (2012) indicated that landslide susceptibility could be influenced in different 363 ways by slope shape and terrain morphology. Curvature represents one of the basic terrain 364

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

variables and utilized in various geomorphometrical analyses (Evans, 1979). For example, plan 365 curvature has a direct impact on the convergence and dispersion of surface runoffs (Nasiri 366 Aghdam et al., 2016). Whereas, the profile curvature affects the material deposition by managing 367 the acceleration or deceleration of these materials on slope (Xiao et al., 2019). In the current 368 work, plan and profile curvatures were extracted from the DEM using ArcGIS 10.2. The plan 369 and profile curvatures were within the ranges of –16.0 to 12.88 and –16.0 to 20.0, respectively 370 (Fig. 4j, k). 371 3.3.11. Landuse/landcover 372 In the study area, human activities such as expansion of urban areas and infrastructure 373 construction are playing significant role in the triggering of landslides. These engineering 374 activities require cutting or excavating slopes for highways and building construction, leading to 375 altering the original geological conditions and subsequently disturb the slope's stability (Xiao et 376 al., 2019). Based on the supervised classification technique of Landsat OLI images (acquired in 377 2016), the study area was classified into five landuse/landcover types including water bodies, 378 buildup areas (residential areas), dense vegetation, rock outcrop, and sparse vegetation (Fig. 4l). 379 3.4. Modelling approach using machine learning techniques

Recently, machine learning techniques (MLTs) becomes a cornerstone in solving spatial 381 modeling problems in the domain of natural hazards (e.g., landslide susceptibility assessment) 382 (Shirzadi et al., 2018; Zhou et al., 2018; Ghorbanzadeh et al., 2019; Sevgen et al., 2019). 383 Marjanović et al. (2011) suggested that MLTs should be utilized as a supportive tool, rather than 384 attempting to replace human expertise. Goetz et al. (2015) recommended that to produce new 385 accurate modeling that could be used by decision makers, Earth scientists and planners, MLTs 386 should be incorporated with the processing capabilities of GIS and field dataset (e. g., historical 387 landslide inventory and geo-environmental factors). The machine learning techniques have some 388 advantages including, (a) their ability to adjust its internal structure to the existing landslide data, 389 (b) their capabilities in extracting knowledge from huge databases in an automatic way, and (c) 390 they have ability to build classification (predicting categorical predictor factors) and regression 391 (predicting continuous dependent factors) in order to provide an accurate landslide model, their 392 models are more cost efficient and rapid than conventional models and can be extended to large 393 area analysis (Felicísimo et al., 2013; Kavzoglu et al., 2019). In the current study, seven 394 advanced machine learning techniques that vary in their degree of complexity were applied to 395 evaluate their efficacy in landslide susceptibility mapping. They include SVM, RF, MARS, 396 ANN, QDA, LDA, and NB. 397 3.4.1. SVM 398 It is a supervised learning method derived from statistical learning theory and the structural risk 399 minimization principle (Vapnik, 1998; Lee et al., 2017). This technique can be used for both 400 classification and regression (Vapnik, 1995; Christianini and Shawe-Taylor, 2000). SVM deals 401 with the binary classification model (Belousov et al., 2016). The SVM algorithm uses the 402 training data to generate a separating hyper-plane in the initial space of coordinates between two distinct categories. The aim of mentioned classification approach is not simply to separate the 404 two classes, but to maximize the margin between them (Yao et al., 2008; Xu et al., 2012). 405 Kanevski et al. (2009) showed that hyper-plane with a large margin should be more resistant to 406 noise and possess better generalization than a hyper-plane with a small margin. It uses kernel 407 functions to map the initial input space into a high-dimensional feature space where the points 408 become more linearly separable (Abe, 2010; Chang and Lin, 2001). Detailed illustrations of 409 SVM modeling were examined by many authors (e.g., Samui, 2008; Yao et al., 2008; Xu et al., 410 2012). To produce a successful SVM training and classification accuracy, it is required to choice 411 of the adequate kernel function (Damaševičius, 2011). Four types of kernel function groups that 412 are commonly used in SVM including linear kernel (LN), polynomial kernel (PL), radial basis 413 function (RBF) kernel, and sigmoid kernel (SIG). In this study, the RBF was used for modeling. 414 The "Kernlab" package (Karatzoglou et al., 2016) was used in R 3.0.2 for landslide susceptibility 415 mapping. 416 3.4.2. RF 417 The RF is an ensemble learning method generating many classification trees that are aggregated 418 to compute a classification (Breiman et al., 1984; Breiman, 2001; Calle and Urrea, 2010; 419 Micheletti et al., 2014). Hansen and Salamon (1990) indicated that an ensemble of classification 420 trees is more accurate than any of its individual members. One of the main advantages of RF is 421 the resistant to over training and growing a large number of random forest trees where it does not 422 create a risk of over-fitting (e.g. each tree is a completely independent random experiment). The 423 RF algorithm data does not need to be rescaled,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

transformed, or modified. It has resistance to 424 outliers in predictors and automatically handles the missing values (Breiman and Cutler, 2004). 425 In this study, landslide susceptibility modeled using ”randomForest” package (Briman and 426 Cutler, 2015) in R 3.0.2 software.

IV. RESULTS AND DISCUSSION

Documentation for Landslide Prediction:
We have a dataset of 780 rows and 21 columns or 21 features to predict the occurence of a landslide. So, the first thing we need to do is clean the dataset with only the features that are required for our prediction model.
After dropping or removing the unnecessary features, we end up with six of them:
Month
Year
Place
Rainfall
Mean Sea Level (M)
Landslide
We Import the Pandas Library which is used for data manipulation and analysis.
Now, we load and read the data.

```
df = pd.read_csv("landslide.csv")
```

A quick way to inspect the data frame is to show the first few lines with the head method:

```
df.head()
```

We should check for any null values that are present in the dataset by

```
In [228]: df.isnull().sum()

Month              0
Year               0
Place              0
Rainfall           0
Mean Sea Level(M)  0
Landslide          0
dtype: int64
```

The dataset contains both numerical and categorical data. Numerical values take continuous values, for example “Mean sea Level (M) Categorical values can have a finite number of values, for example "Place”.
We can check the number of samples and the number of columns available in the dataset and we can compute the number of features by counting the number of columns and subtract 1, since one of the columns is the target:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

```
In [221]: print(f"The dataset contains {df.shape[0]} samples and "
                f"{df.shape[1]} columns")

          The dataset contains 780 samples and 6 columns

In [220]: print(f"The dataset contains {df.shape[1] - 1} features.")

          The dataset contains 5 features.
```

Next we import the LabelEncoder from sklearn.preprocessing to fit the data such that we end up only with numerical values. LabelEncoder is usually used to convert non-numerical values to numerical values.

In [233]: df

	Month	Year	Place	Rainfall	Mean Sea Level(M)	Landslide
0	4	0	3	0.0	179	0
1	3	0	3	0.0	179	0
2	7	0	3	53.2	179	0
3	0	0	3	0.0	179	0
4	8	0	3	243.4	179	0
...	...	...	...	...	...	...
775	1	12	1	266.4	168	0
776	11	12	1	238.4	168	0
777	10	12	1	143.3	168	0
778	9	12	1	24.8	168	0
779	2	12	1	64.6	168	0

780 rows × 6 columns

We see that this CSV file contains all information: the target that we would like to predict (i.e. “Landslide") and the data that we want to use to train our predictive model (i.e. the remaining columns). The first step is to separate columns to get on one side the target (i.e., the y value) and on the other side the data (i.e., the x value).

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

```
In [234]: x=df.iloc[:,:-1]
          x
```

	Month	Year	Place	Rainfall	Mean Sea Level(M)
0	4	0	3	0.0	179
1	3	0	3	0.0	179
2	7	0	3	53.2	179
3	0	0	3	0.0	179
4	8	0	3	243.4	179
...	...	...	...	...	...
775	1	12	1	266.4	168
776	11	12	1	238.4	168
777	10	12	1	143.3	168
778	9	12	1	24.8	168
779	2	12	1	64.6	168

780 rows × 5 columns

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

```
In [235]: y=df["Landslide"]
          y

0      0
1      0
2      0
3      0
4      0
      ..
775    0
776    0
777    0
778    0
779    0
Name: Landslide, Length: 780, dtype: int64
```

When building a machine learning model, it is important to evaluate the trained model on data that was not used to fit it, as **generalization** is more than memorization (meaning we want a rule that generalizes to new data, without comparing to data we memorized). It is harder to conclude on never-seen instances than on already seen ones.

Correct evaluation is easily done by leaving out a subset of the data when training the model and using it afterwards for model evaluation. The data used to fit a model is called training data while the data used to assess a model is called testing data.

We can load more data, which was actually left-out from the original data set.

Now, We import the Logistic Regressor from sklearn. The fit method is called to train the model from the input (features) and target data.

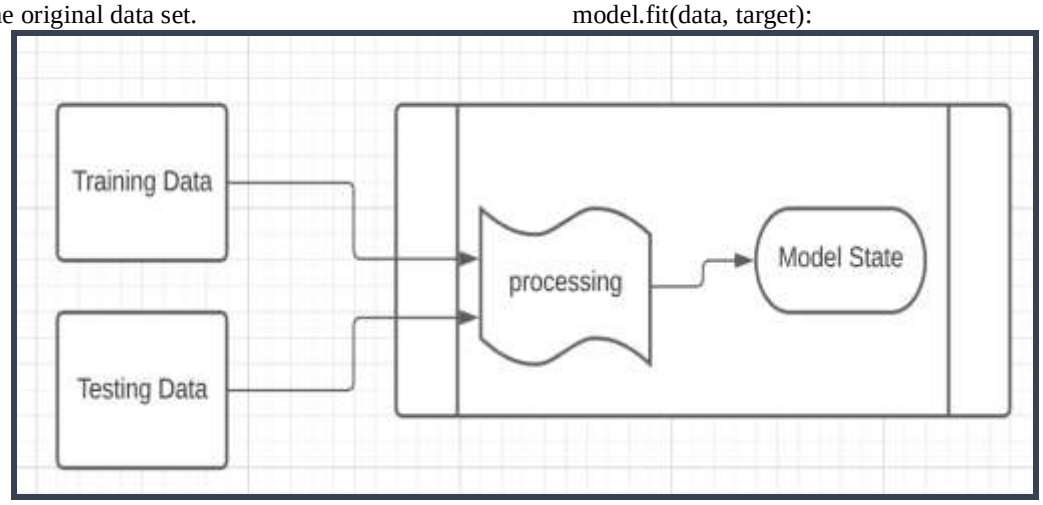
The method fit is composed of two elements:

(i)a**learning algorithm**(ii) some **model states**.

The learning algorithm takes the training data and training target as input and sets the model states. These model states will be used later to either predict (for classifiers and regressors) or transform data (for transformers).

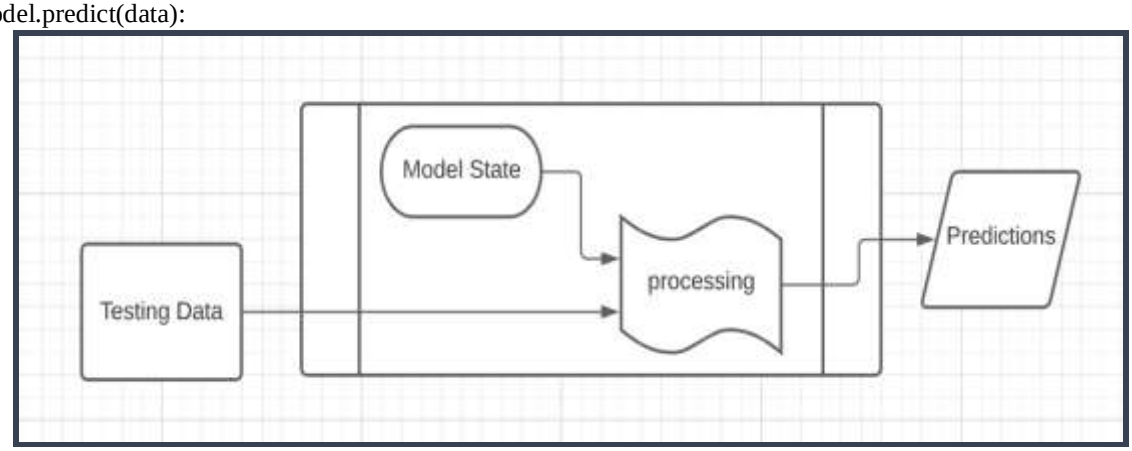
Both the learning algorithm and the type of model states are specific to each type of model.

model.fit(data, target):



International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

model.predict(data):



To predict, a model uses a **prediction function** that will use the input data together with the model states. As for the learning algorithm and the model states, the prediction function is specific for each type of model.

Now, We check the accuracy score for our model by importing accuracy_score from sklearn.metrics:

```
In [239]: from sklearn.metrics import accuracy_score
          accuracy=accuracy_score(y_test,y_pred)
          accuracy

          0.9615384615384616
```

We got an accuracy score of 0.96 which is a pretty good score for the model.

A confusion matrix is **a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.** We can get it by importing it from sklearn.metrics.

```
In [240]: from sklearn.metrics import confusion_matrix
          confusion=confusion_matrix(y_test,y_pred)
          confusion

          array([[148,   3],
                 [  3,   2]], dtype=int64)
```

We can now predict the landslide by inputting the values of Month, year, place, rainfall and mean sea level(M) in:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

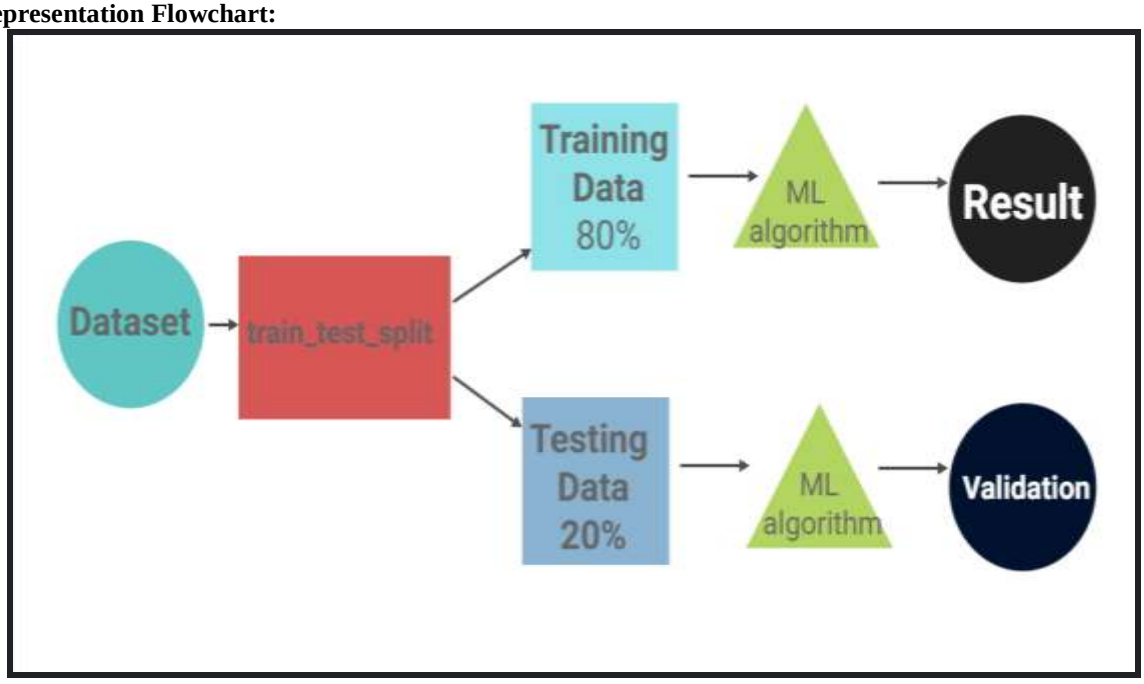
```
In [242]: lr.predict([[4,12,1,56,0]])

          array([0], dtype=int64)
```

We input the values in the order
lr.predict([[Month, year, place, rainfall, mean sea level]])
and we either get 0 or 1 in the output inside
array([ ]
If we get 0, it means that there is no landslide occurrence whereas if we get 1, then there is a landslide occurrence.
We got 0 which means there is no landslide occurrence there.
Similarly we can create the model using Decision tree classifier, regressor and support vector machine models.
The accuracy rate obtained by each models are given below:

- Logistic Regression: 0.96
- Decision Tree Classifier: 0.96
- Support Vector Machine: 0.95

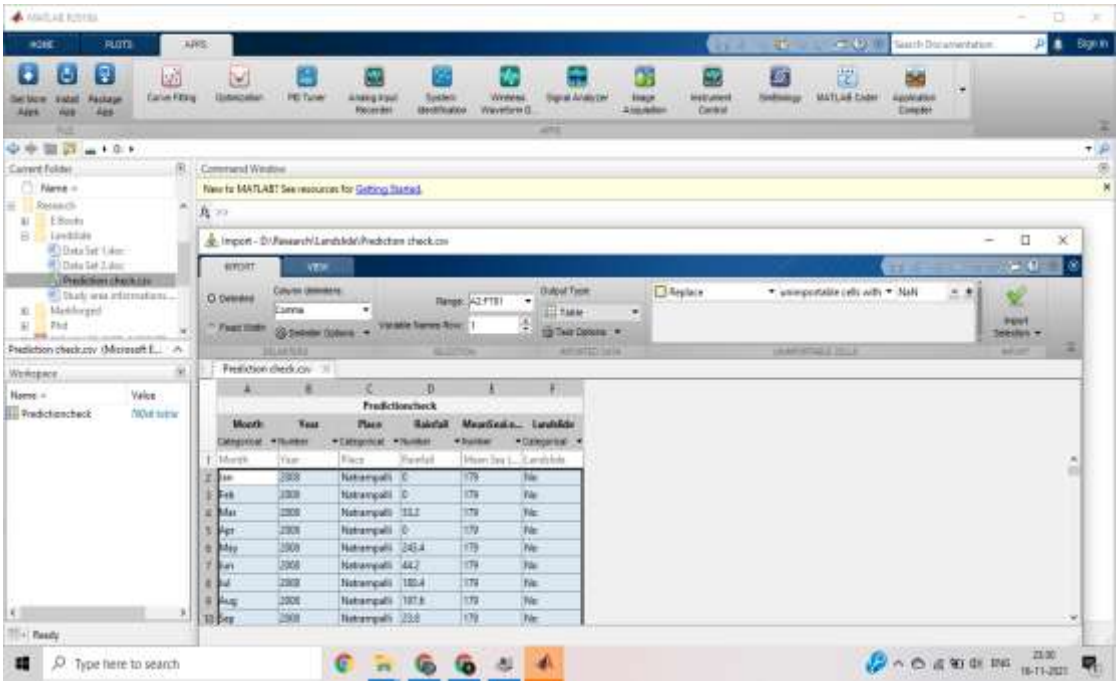
We collect the required data and the check for the null values first using the isnull() function. Then we import the labelencoder library which converts all the texts or string data/values into numerical values. Now, we split the data using the train_test_split method with 20% of the data for testing and the rest 80% of the data for training. Now, we import the necessary library packages from sklearn like the logisticRegression, decision tree, random forest, and suppose vector machine (SVM), etc and fit the data according to the input values for each algorithm using fit function and then predict the output using the inbuilt function called as predict in the sklearn library.

Representation Flowchart:

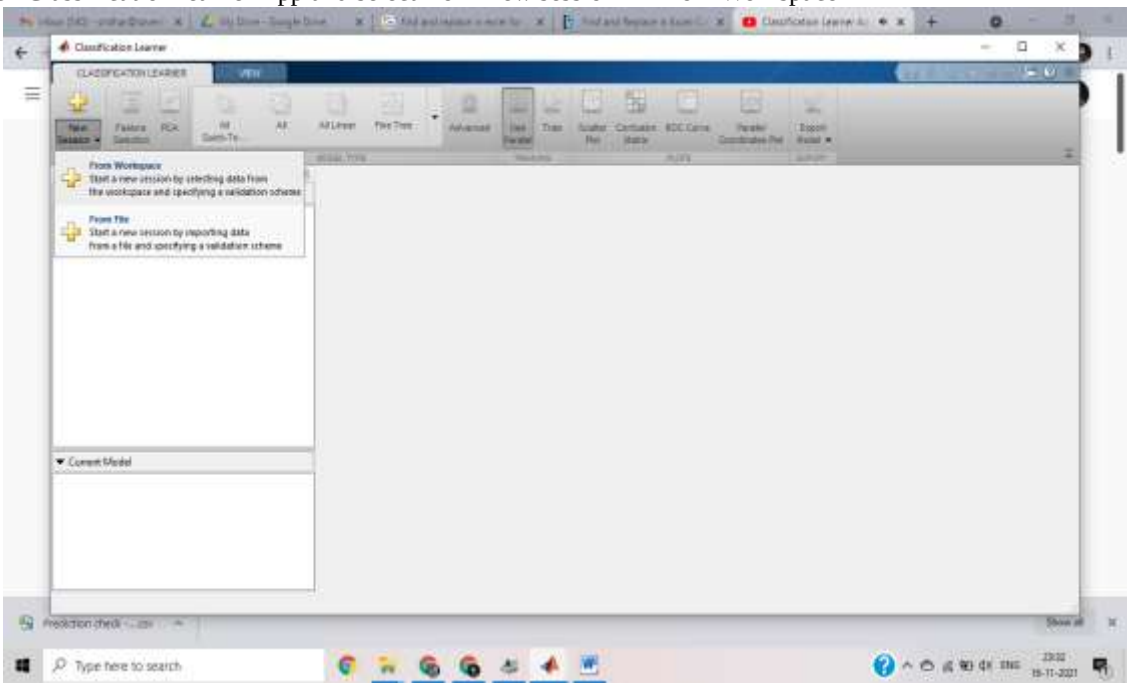


Importing Dataset into workspace from directory

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

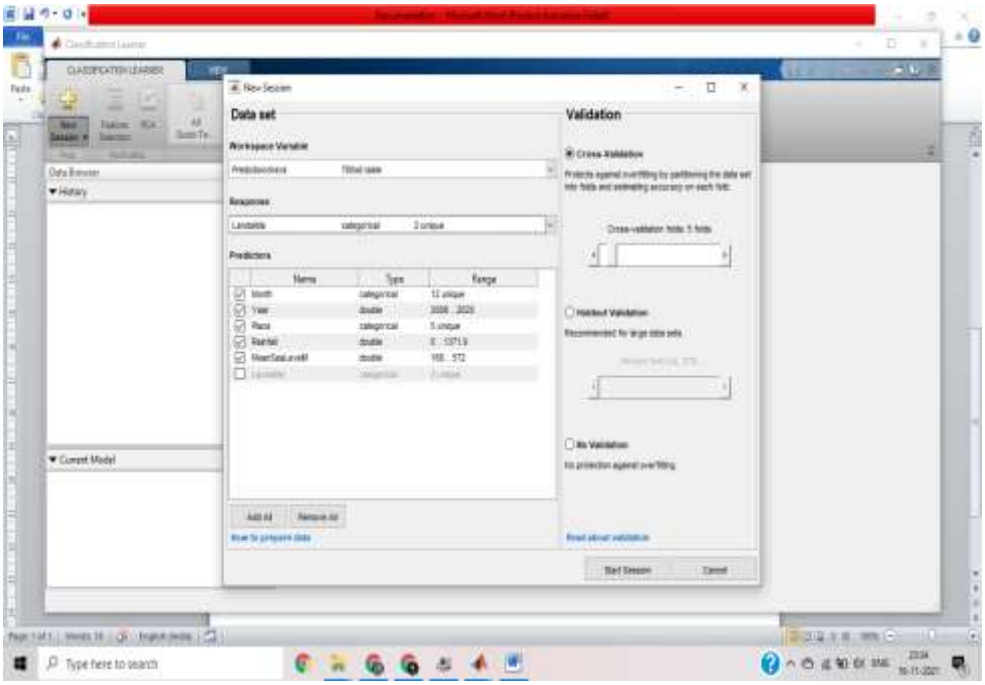


Open Classification learner App and select from New session → From Workspace

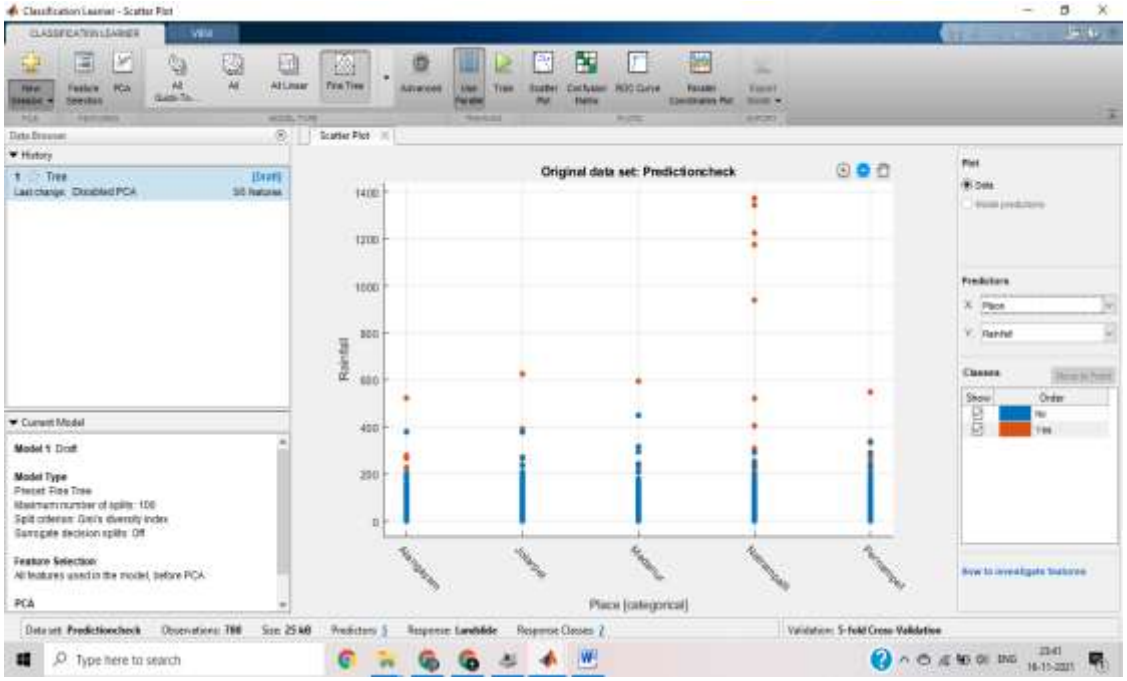


Set the response as “Landslide” and other entities as predictors.
For validation select “Cross- Validation” with 5 folds since data in very less & then click “start session”

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252

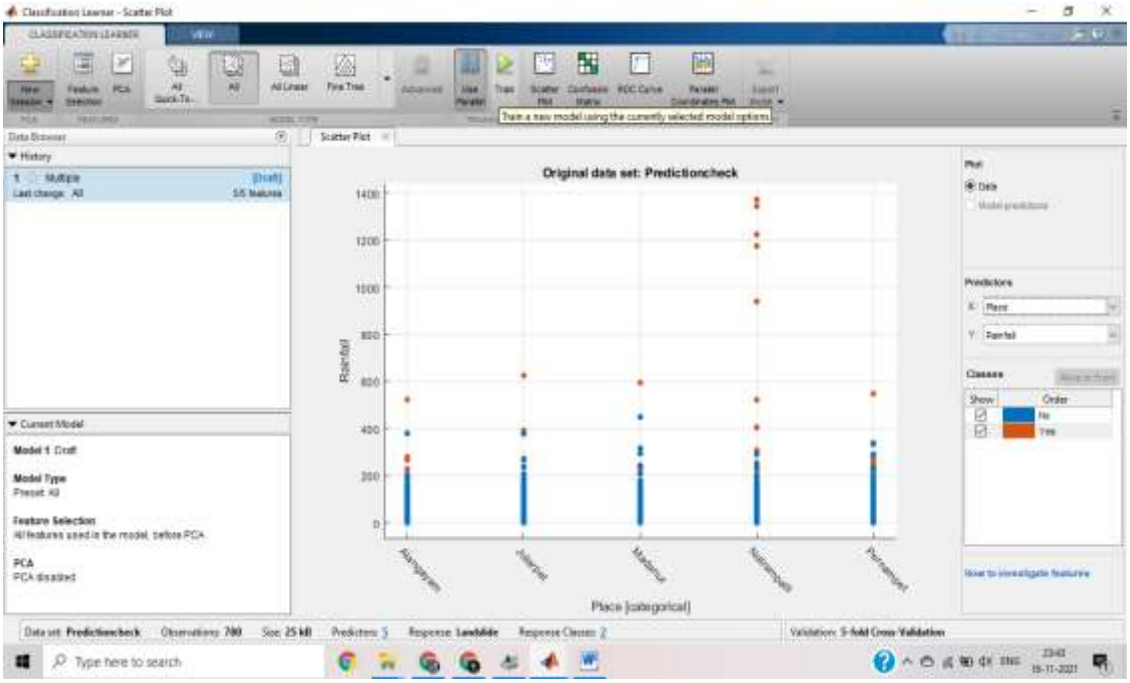


Checking the dataset

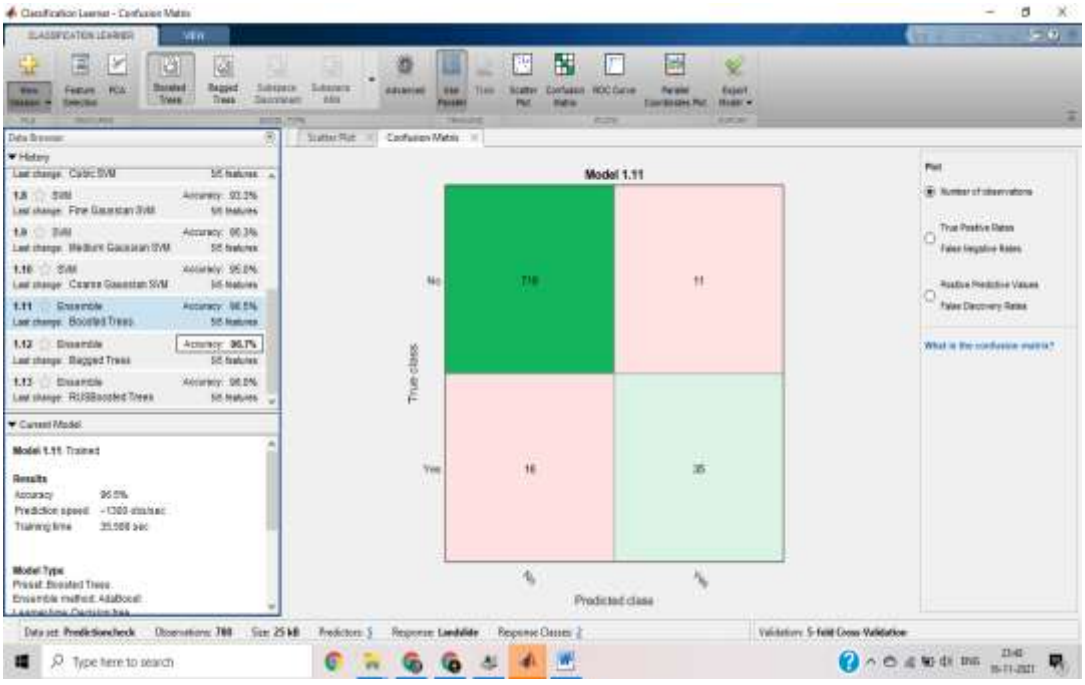


We are unsure about which model will give with high prediction accuracy. So training the model in all classifiers. Select “ALL” and click “Train”.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 1458-1477 www.ijaem.net ISSN: 2395-5252



Almost all the models have good accuracy. Selecting two models for accuracy check with test data.
Models to be exported are
1) Ensemble – Boosted trees 96.5%
2) SVM – Medium Gaussian SVM 95.3%



Hence, The following algorithms gave the below accuracy score for the Landslide prediction model:

- Logistic Regression: 0.96
- Decision Tree Classifier: 0.96
- Support Vector Machine: 0.95

V. CONCLUSIONS

A survey of machine learning algorithms applied in landslidesprevention has been presented in this paper, whichfocuses on (1) landslides detection based on images, (2)landslides susceptibility assessment, and (3) the

developmentof landslide warning systems. The survey shows thatmachine learning methods have been widely used inlandslides prevention and can achieve satisfactory perfor-mance. However, there are still several challenges andlimitations. First, professional knowledge is needed, whichcan facilitate the selection of more appropriate variablesand datasets when facing increasingly complex and massivedata. Second, interpretability is also a critical componentin landslides prevention. The majority ofestablished scientific theories on landslide occurrencemechanisms struggle to explain the results of machinelearning models. Analyses involving machine learningresults should be interpreted in combination with landslidemechanisms. Therefore, a potential research trend is tocombine data-driven machine learning with expert knowledgeof landslides. Gradually increasing the application ofmachine learning in landslides prevention will enablebenefits for both the machine learning and landslideresearch domains.

VI. AUTHOR CONTRIBUTIONS

S G: conceived, designed, conducted, analysis interpretation of data and drafted the manuscript, conducted the literature search and drafted the manuscript.

R S: drafted the manuscript and was involved in the analysis interpretation of data.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2021.111238.

REFERENCE

[1]. **S.** Arya, G. Vennila and T. Subramani, Spatial and seasonal variation of groundwater levels in Vattamalaikarai River basin, Tamil Nadu, India - A study using GIS and GPS. Indian J. Geo-Mar. Sci.,47(9) (2018) 1749-1753.

[2]. K.R. Karanth, Groundwater assessment development and management. Tata McGraw hill, New Delhi, 1987.

[3]. J. Krishnamurthy, A.N. Mani, V. Jayaram and M. Manivel, Groundwater Resources Development in Hard Rock Terrain: An Approach Using Remote Sensing and GIS Techniques. Int. J. Appl. Earth Obs. Geoinf., 2(1) (2000) 204-215.

[4]. Binay Kumar and Uday Kumar, Integrated approach using Remote Sensing and GIS techniques for mapping of groundwater prospects in Lower Sanjai Watershed, Jharkhand. Int. j. geomat. geosci., 1(3) (2010) 587-598.

[5]. A.K. Saraf and P.R. Choudhury, Integrated remote sensing and GIS for ground water exploration and identification of artificial recharge sites. Int. J. Remote Sens., 19(10) (1998) 1825-1841.

[6]. G. Venkatesan and P. Raj Chandar, Possibility studies and parameter finding for interlinking of Thamirabarani and Vaigai Rivers in Tamil Nadu, India. Int. J. Earth Sci. Eng,1(1) (2012) 16–26.

[7]. G. Venkatesan and T Subramani, Case Study on Environmental Impact Due To Industrial Waste Water In Vellore District, Tamil Nadu, India Using Geospatial Techniques. Middle East J. Sci. Res., 24(1) 2016a 152-159.

[8]. C.T Anuradha and S. Prabhavathy, Water resources management for Virudhunagar district using Remote Sensing and GIS. Int. J. Earth Sci. Eng., 3(1) (2010) 55-61.

[9]. J. Krishnamurthy, P. Manavalan and V. Saivasan, Application of digital enhancement techniques for groundwater exploration in a hard rock terrain. Int. J. Remote Sens., 13(15) (1992) 2925-2942.

[10]. Y. Srinivasa Rao and D.K. Jugran, Delineation of groundwater potential zones and zones of groundwater quality suitable for domestic purpose using remote sensing and GIS. Hydrol. Sci. J. HYDROLOG. SCI. J., 48(5) (2003) 821-833.

[11]. S. Venkateswaran, M. Vijay Prabhu and S. Karuppannan, Delineation of Groundwater Potential Zones Using Geophysical and GIS Techniques in the Sarabanga Sub Basin, Cauvery River, Tamil Nadu, India. Int. j. curr. res. acad. rev., 2(1) (2014) 58-75.

[12]. N. Thilagavathi, T. Subramani, M. Suresh and D. Karunanidhi, Mapping of groundwater potential zones in salem chalk hills, Tamil Nadu, India, using Remotes Sensing and GIS Techniques. Environ. Monit. Assess., 187(2) (2015) 164-181.
[13]. T. Subramani, B. Savithri and L. Elango, Computation of groundwater resources and recharge in Chithar River basin, South India. Environ. Monit. Assess., 185(1) (2013) 983-994.
[14]. D. Sivakumar, V. Balasundaram, G. Venkatesan and S.P. Saravanan, Effect of tamarind kernel powder for treating dairy industry wastewater, Pollut. Res., 33 (2014) 519-523.
[15]. G. Venkatesan and T. Subramani, Environmental degradation due to the Industrial Wastewater discharge in Vellore District, Tamil Nadu, India. Indian J. Geo-Mar. Sci., 47(11) (2018) 2255-2259.
[16]. G. Venkatesan, T. Subramani, U. Sathya and D. Priyadarsi Roy, Seasonal changes in groundwater composition in an industrial center of south India and quality evaluation for consumption and health risk using geospatial methods. Chem. Erde.,80(4) (2020) 125651.
[17]. Groundwater perspectives – A profile of Vellore District Tamil Nadu Public Works Department (PWD): Government of Tamil Nadu, India, 2011.
[18]. Prickett Lennquist. Selected digital computer techniques for groundwater resources evaluation. Illinois State Water Survey, Bull., 55(2) (1971).
[19]. G. Venkatesan and T. Subramani, Reduction of Hexavalent Chromium to Trivalent Chromium from Tannery Effluent using Bacterial Biomass. Indian J. Geo-Mar. Sci., 48(4) (2019) 528-534.
[20]. G. Venkatesan, T. Subramani, D. Karunanidhi, U. Sathya and Peiyue Li, Impact of precipitation disparity on groundwater fluctuation in a semi-arid region (Vellore district) of southern India using geospatial techniques. Environ. Sci. Pollut. Res. Int., 28(15) 2020 18552.
[21]. S. Anandakumar, T. Subramani and L. Elango, Spatial variation and seasonal behaviour of precipitation pattern in Lower Bhavani River basin, Tamilnadu, India. Ecoscan,2(1) (2008) 17-24.
[22]. M.P. Sharma, Anukaran kujur and Udayan Sharma, Identification of groundwater prospecting zones using Remote Sensing and GIS techniques in and around Gola block, Ramgargh district, Jharkhand, India. Int. j. sci. eng. res., 3(3) (2012) 01-06.
[23]. S. Srinivasa Vittala, S. Govindaiah and H. Honne Gowda, Digital Elevation Model [DEM] for identification of Groundwater prospective zones. J. Indian Soc. Remote. Sens.,34(3) (2006) 319-324.
[24]. C. Mohanty C and S.C. Behrera, Integrated Remote sensing and GIS study for Hydrogeomorphological mapping and Delination of groundwater potential zones in Khallikote Block, Ganjam District, Orissa. J. Indian Soc. Remote. Sens., 38(2) (2010) 345-354.
[25]. G. Venkatesan G, R. Aishwaryya, A.S. Renjinny and M. Pavithra, Surface & Groundwater Management A Remote Sensing and GIS based. International Journal for Scientific Research and Development., 1(2) (2014) 158-162.
[26]. P.P. Schot and J. Vander Wal, Human impact on regional groundwater composition through intervention in natural flow pattern and changes in land use. J. Hydrol., 134(1-4) (1992) 297-313.
[27]. F.G. Bell. S.E.T. Bullock, T.F.J. Halbich and P. Lindsey, Environmental impacts associated with an abandoned mine in the Witbank Coalfield, South Africa. Int. J. Coal Geol., 45(1) (2001) 195-216.
[28]. M. Nagarajan and Sujit Singh, Assessment of Groundwater Potential Zones using GIS techniques. J. Indian Soc. Remote. Sens., 37(1) (2009) 69-77.
[29]. P. Rasmussen, Monitoring shallow groundwater quality in agricultural watersheds in Denmark. Environ. Geol., 27(4) (1996) 309-319.

[30]. G. Venkatesan, T. Subramani, U. Sathya and D. Karunanidhi, Evaluation of chromium in vegetables and groundwater aptness for crops from an industrial (leather tanning) sector of South India. Environ. Geochem. Health., 43(2) (2020) 995–1008.
[31]. B. Pradhan, Groundwater potential zonation for basaltic water- sheds using satellite Remote Sensing data and GIS techniques. Central Eur. J. Geosci., 1(1) (2009) 20-129.
[32]. M. Samake, Z. Tang, W. Hlain, N. Mbue and K. Kasereka, Assessment of Groundwater Pollution Potential of the Datong basin, Northern China. J. Sustain. Dev., 3(2) (2010) 140–152.
[33]. R.E. Horton, Erosional development of streams and their drainage basins: hydrophysical approach to quantitative morphology. Geol. Soc. Am. Bull., 56(3) (1945) 275–370.
[34]. Biswajeet Pradhan, Groundwater potential zonation for basaltic watersheds using satellite remote sensing data and GIS techniques. Cent. Eur. J. Geosci., 1(1) (2009) 120-129.
[35]. E. Bocanegra, O.M.Q. Londono, D.E. Martinez and A. Romanelli, Quantification of the water balance and hydrogeological processes of groundwater lake interactions in the Pampa Plain, Argentina. Environ. Earth Sci.,68(1) (2013) 2347-2357.
[36]. R.K Prasad, N.C. Mondal, P. Banerjee, M. Nandakumar and V.S. Singh, Deciphering potential groundwater zones in hard rock through the application of GIS. Environ. Geol., 55(3) (2008) 467-472.
[37]. G. Venkatesan and T. Subramani, Parameter Finding for Case Study of Environmental Degradation due to Industrial Pollution in Vellore, Tamil Nadu, India Using Remote Sensing and GIS. IJESIT, 1(1) (2016) 1-7.
[38]. N. J. Raju, T.V.K. Reddy, B. Kotaiah and P.T. Nayudu, Hydrogeomorphology of the upper Gunjanaeru river basin, Cuddapah district, Andhra Pradesh using remote sensing techniques. Journal of Applied Hydrology, VIII (1-4) (1995) 99-104.
[39]. N. J. Raju, T.V.K. Reddy and P. Munirathnam, Subsurface dams to harvest rainwater – a case of Swarnamukhi River basin, Southern India. Hydrogeology Journal, Springer 14(4) (2006) 526-531.
[40]. N. Janardhana Raju, T. V. K. Reddy, P. Muniratnam, Wolfgang Gossel and Peter Wycisk, Manged Aquifer Recharge (MAR) by the construction of subsurface dams in the semi-arid regions: a case study of the Kalangi river basin, Andhra Pradesh, India. Journal of Geological Society of India, 82(6) (2013) 657-665.
[41]. D. Karunanidhi, P. Aravinthasamy, T. Subramani, Deepak Kumar and G. Venkatesan, Chromium contamination in groundwater and Sobol sensitivity model based human health risk evaluation from leather tanning industrial region of South India. Environ Res., 199 (2021) 111238.