IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 January 2023, accepted 2 February 2023, date of publication 22 February 2023, date of current version 21 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3247499

RESEARCH ARTICLE

Monitor the Strength Status of Buildings Using Hybrid Machine Learning Technique

VISHNU VARDHANA RAO M.[1], APARNA CHAPARALA[2], PRAPHULA KUMAR JAIN[3], HANEN KARAMTI[4], AND WALID KARAMTI[5,6]

[1]Department of Computer Science and Engineering, Vignan's Institute of Management and Technology for Women, Kondapur, Ghatkesar, Medchal, Telangana 501301, India
[2]Department of Computer Science and Engineering, R. V. R. & J. C. College of Engineering, Guntur, Andhra Pradesh 522019, India
[3]Department of Computer Engineering and Applications, GLA University, Mathura, Uttar Pradesh 281406, India
[4]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[5]Department of Computer Science, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia
[6]Data Engineering and Semantics Research Unit, Faculty of Sciences of Sfax, University of Sfax, Sfax 3029, Tunisia

Corresponding author: Praphula Kumar Jain (praphulajn1@gmail.com)

This work was supported by the Princess Nourah Bint Abdulrahman University Researchers Supporting through Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia, under Project PNURSP2023R192.

ABSTRACT Standard inspections of buildings are not always possible because of human flaws in prediction. Hence, we need more stable, scalable, and efficient automated processes. Structure Health Monitoring (SHM) is one of the automation systems for forecasting potential losses in building structures. This article suggested how to monitor the strength status of buildings by using Hybrid Machine Learning Technique (HMLT). HMLT contains two-hybrid procedures. One for identifying the most significant features in Dataset using Hybrid Feature Selection Method (HFSM). HFSM uses the combined features of Mutual information (MI) and Rough Set Theory (RST) for feature selection. Another method is optimized classifiers such as Support Vector Machine (SVM) and Artificial Neural Networks (ANN) are used for the classification and predicting the accuracy i.e. predicting the strength status of buildings. Now the proposed method is applied on Earthquake Damage Dataset (Gorkha Earthquake in April 2015). Training and 10- fold cross-validation procedure pragmatic to features. Then the performance of proposed method has been evaluated using the F1-score and accuracy metrics and get 91% and 92% respectively. Finally, the result analysis demonstrates the importance of the proposed approach in predicting the status of the building strength.

INDEX TERMS KNN: K-nearest neighbors, RF: random forest, GBM: gradient boosted machines, SVM: support vector machine, ANN: artificial neural networks, HMLT: hybrid machine learning technique, SHM: structure health monitoring system, HFSM: hybrid future selection methodology.

I. INTRODUCTION

The definition of an Earthquake is the vibration of the earth or the pulsating of the ground. Natural, human-made, and artificial/induced seismicity are some types of earthquakes. Natural Disasters generates the damages in civil-infra-structures. Rathnaweera et al. [1] and Cremen et al. [2] presented the percentage of earthquakes that occurred in natural is 0-89%, and the remaining 0-9% belongs to human-made hazards/other issues. In recent years, the damage rate in the buildings and their structures are increased rapidly because of environmental changes and human errors. Early detection of disasters mitigates the damage rate as well as death rate.

The abnormal changes in Animal Behavior [3], Temperature [4], [5], Lunglin-1976 and Przhevalsk -Russia 1970, Water Levels [6], [7], Earthquake in 2016 at Kumamoto), Velocity of P-wave (V_P) and Velocity of S-wave (V_S), use as instances for early detection of disasters. "Damage" can be defined as the changes in the properties of the building structures, boundary environments and system connectivity such as geo-metric and/or material. It shows the unsympathetically impact on the performance of the model. In the opinion

The associate editor coordinating the review of this manuscript and approving it for publication was Paolo Crippa.

TABLE 1. Building damage levels for earthquakes.

EVENT-YEAR	ORGANIZER	TYPE OF BUILDINGS	NUMBER OF DAMAGE LEVELS	NAMES	USED NAME
Mexico earthquake-1985	Architectural Institute of Japan	All types**	6	"No-damage", "Negligible-damage", "Slight - damage", "Moderate-damage", "Major-damage", & "Collapse"	Ranks
Hyogo-ken Nanbu earthquake-1995	Japanese PMO(Prime-Minister's-Office)	wooden framed buildings	4	"No-damage", "Moderate-damage", "Heavy-damage", and "Major-damage".	Grade
Indian Ocean Tsunami at Sri Lanka-2004	European Macro Seismic scale (EMS)-98	All types**	4	"No/slight damage", ``Moderate damage", ``Heavy damage' 'and ``Complete damage".	Grade
Mexico earthquake-1985	Ministry of Land, Infrastructure, Transport and Tourism (MLIT)	All types**	6	"Nodamage","Minor","Moderate","Major","Complete","Collapsed","Washed Away"	Ranks
Mexico earthquake-1985	Federal Emergency Management Agency (FEMA)	All types**	4	"Slight", "Moderate", "Extensive", "Complete".	Grade
Mexico earthquake-1985	European Macro Seismic Scale (EMS)-983	All types**	6	"No damage", "Slight damage", "Moderate Damage", "Heavy damage", "Very heavy", "Destruction".	Grade

of Yao et al. [8] "damage" is defined as the strength of the buildings decreasing due to the impact of load, human, and environmental errors. If the change is less than the threshold limit then structures are considered as damage-free; otherwise, damage is observed in the structure of the building. Kirkegaard et al. [9] defines the different Damage Levels (LE),

LE-1: Assesses the presence or occurrences of damage in the building structures.

LE-2: Estimates the place or location of the damages in the structures.

LE-3: Speaks about the damage impact or severity in the structures.

LE-4: It declares that the leftover the provision of the structure (lifetime of structures).

In 2011, triggered the Earthquake with the nine-magnitude at the northern place of Japan. The name is called GEJET-Grate East Japan Earthquake, affecting the number of buildings and showing a very dangerous impact on humans [10], [11], [12], [13], [14], [15]. Table-1 shows that different types of damage levels, names defined by the organizers followed by the event year, type of buildings, and used name [18]).

The below Table-2 shows that Grade Descriptions with respect to the grade numbers, sample diagrams with different shapes in Building structures. If building structures are experienced the natural disaster like Earthquake. Each structure is defined as the name of the damage type based on the above standards. SHM is one of the most robust automatic predicting framework for finding building strength status based on the earthquake damage dataset. In this article, our contribution is Hybrid Machine Learning Technique (HMLT), measuring the strength status of buildings using Two-Hybrid methods, one for Feature Reduction Selection Method (HFSM) and another for Classification and Damage Prediction.

The organization of this article is as follows: In section-II, discusses the reasons for why we take this problem (Related work). Dataset Visualization and what is the impact of Damage Grade with other features in the Earthquake Damage Dataset in Section-III. In Section-IV develop the Algorithms for implementing the Two Hybrid Models and Proposed Methodologies are explained in section-V. Finally, the result analysis is shown in Section-VI. In section-VII presents the conclusions and future scope.

II. RELATED WORK

Due to Natural disasters the abnormal changes are done in strength status of buildings. So SHM considered as the useful automation tool for estimating the health status of the structures and evaluate the building damage levels (LE). Machine Learning-ML, Artificial Intelligence-AI, and Statistical Techniques such as ANN, genetic algorithm-GA, SVM, and Principal component analysis-PCA methods are used for estimating structural, damage levels in civil structures. Debnath et al. [16] analysed the impact of earthquakes in India by using ML classifiers and develop the vibration-based-SHM model for forecasting strength status of the buildings based on the damage datasets. Train the model with seven different ML Classifiers and also completed the analysis with six different datasets and regions in India then the results says that the investigated method is more suitable for forecasting strength status of the buildings.

Lu et al. [17], done a simple re-review on civil engineering problems, the reviews done based on the AI. Saengtabtim et al. [18] completed the Predictive Analysis on the 2011 Great East Japan Tsunami for the Building Damage Using DT (Decision Tree).

TABLE 2. Grade descriptions.

Example Diagram	Grade Number	Grade Description	Status of Building
	Grade-1	No Damage	No Need to Monitoring
	Grade-2	Moderate Damage	Need Monitoring
	Grade-3	Substantial to Heavy Damage.	Need Monitoring
	Grade-4	Very Heavy Damage	Replace is Need
	Grade-5	Destruction or Collapse	Replace With New Building is Need

Bao et al. [19], Developed the Novel Structural Health methodology based on immeasurable monitoring information. Bao et al. [20], suggested that Machine-Learning-Structural Health Monitoring (MLSHM) for predicting the status of strength of the building.

Adeli et al. [21] revised the work, integration of ANNs with various paradigms, such as GA, fuzzy logic, and wavelet analysis. Meng et al. [22], inveterate or deep-rooted the unassailability of using metaheuristics-optimization problems.

Dave et al. [23], suggested the method for estimating the faults in bearing. The vibration signals are used for predict the bearing faults, for feature selection used the mutual information feature ranking method, and for accurate calculation used ML classifiers. Tenfold cross validation procedures applied to all ranked features and training them. This method estimate the bearing faults, but SVM and ANN gives the accuracy decreases based on the ranked features, 90% and 89 % with 21 and 12 ranked features are observed and training, lowest accuracy witnessed of 50 % with ANN, 62.5 % with SVM and only one ranked feature followed, one ranked feature respectively. When cross-validation is performed then the minimum accuracy observed is 59.3 %, 62.5 % with ANN, SVM, respectively with one, one and thirteen features respectively. The maximum ten-fold cross-validation accuracy achieved is 98.43 % with seven ranked features.

Ahadzadeh et al. [24] recommended a methodology that applied social media data for the earthquake damage assessment at the county, city, and 10km grids scale using Naive Bayes, support vector machine (SVM), and deep learning classification algorithms. Using these methods to classify the messages as damage and non-damage called binary classification. For accumulative the awareness on the post crisis, situation author suggested that multi-classification of message. Using metrics of ML (accuracy, precision, recall, and F-score), classification was evaluated. In the binary classification (multi-class classification), the SVM algorithm performed better in all the indices, gaining 71.22% (90.25%) accuracy, 81.22% (88.58%) F-measure, 79.08% (84.34%) accuracy, 85.62% (93.26%) precision.

Hongfang et al. [25] suggested that hybrid method for feature selection called CCMI (Mutual Information with Correlation Coefficient). CCMI are used to select a feature subset that is highly relevant to the class and has low redundancy between features. Basically MI are used for to measure the relationship between the class-label and features and between features and features (filter out more redundant information). Correlation Coefficient (CC), assesses the degree of redundancy between features and measure the importance of redundant items in the evaluation function and also use the principle of minimization for evaluation criteria. CCMI method are applied on 12 types of datasets and observed that this method gives the better classification accuracy Steppe et al. [26] suggested that hybrid feature selection method called Rough-Set with Mutual Information-RSMI. RSMI is used for selecting the most significant features from the dataset. These features are used for getting better results but the problem is, applied on linear only. Kwak et al. [27] implemented the Mutual Information with Feature Selection MIFS method for feature selection worked on non-linear problems then enhance this to MIFS-U gives the higher accuracy with compared to other feature subset selection methods.

Srivastava et al. [28] implemented the hybrid model called Fuzzy-RST (FRST). The Fuzzy Set Theory contains the t-norms and t-conforms, used for select the optimal features. Srivastava et al. [29] implemented hybrid techniques for classification problems called Rough-SVM. RSVM are used for data classification. Now this method is compared with SVM, ANN and KNN got better accuracy.

Ulrike et al. [30] developed the Hybrid Method called Damage Index Method–DIM for Damage Identification utilising Artificial Neural Networks-ANN. Zuowei et al. [31] defines, an integrated method called RST with ANN used for predicting the damage and this method is enhancement of the MIFS. Many research work included ML techniques for prediction [52], [53], [54], [55].

In my previous works, first completed the comparisons between five different classifiers applied on earthquake-damaged dataset for predicting the damage levels of buildings [32]. Next, the number of features in the dataset is one of the issues to reduce the accuracy. So using hybrid feature reduction method (MI&RST) to get the maximum accuracy at the same time to minimize the time for constructing the model [33]. For getting better results implements, the hybrid method for estimating the damage levels of the buildings called Rough-SVM and Rough-ANN, used for Feature Reduction and accuracy of classifiers. RS-SVM is used to classify the structures, and RS-ANN is used to predict the damage levels. This method applied on earthquake-damaged datasets, got the accuracy of 90% [34]. In the next contribution we implement the hybrid method for both feature reduction (MI+RST), classification and predicting the damage levels in the earthquake damage datasets (RAS-method) [35].

In all my works, uses the one automatic tool called SHM and maintain one working procedure for Damage Predicting as shown in below Figure-1: Figure-1 shows that, the total workflow for damage forecast. This system contains two steps, namely pre-processing step and the Damage predicting step. First, create the dataset that gathered desirable information from sensors. For identification of most significant features from the dataset use various feature reduction methods (Pre-Processing step). ML classifiers are used for classification, with our objective predicting the damage level of the buildings.

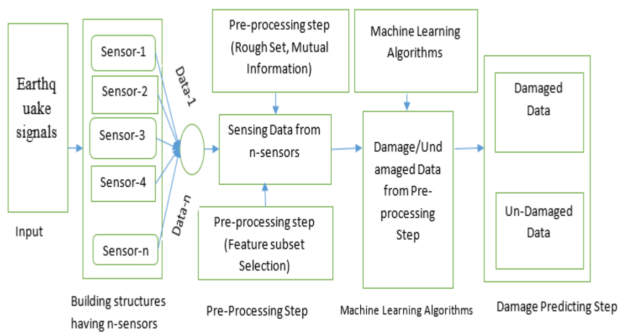


FIGURE 1. Work flow for the damage predicting system.

In the next section discussed about the Dataset Visualization [36], in this paper, implemented the Hybrid Feature Selection Method (HFSM) for choosing better Feature and

Hybrid Pipeline Technique for getting better results, increasing prediction accuracy.

III. VISUALIZATION OF DATASET

In this section, we explain about the Multi-Class Supervised learning problem to predict the damage classification and strength status of buildings from the Earthquake Damage Dataset. Here we considered the Earthquake Damage Dataset with the magnitude of 7.8, occurred at the Gorkha district of Gandaki Pradesh, Nepal in 2015. Almost 0-8999 lives were lost and \$10 billion loss. Next we explain about the Dataset Information.

A. DATASET INFORMATION

Dataset contains 30 features. Dataset contains 30 features. Dataset has 762106 rows and 30 columns of information; one of the fields is our target variable, which is called 'Damage Grade' and it contains the two types of variables called categorical variables and Numerical variables. The categorical variables are District, Land surface type, Foundation type, Roof type, Ground Floor type, other Floor type, position, building shapes, plan configuration and condition in Dataset and Numerical variables are number of room, age of the building, height of the buildings and attributes starts with "has-prefix" in Dataset before, after Earthquake.

B. DATA REFERENCE

Dataset has 762106 rows and 30 columns of information; one of the fields is our target variable, which is called 'damage_grade'. The target variable has five classes, labelled 'Grade 1':'Grade 5', which each represent the different scale of damage sustained to the building.

The below Table-3 shows that most of the data is fully intact; there are a handful of columns whose fields are not fully populated. As there are only a small number of missing entries, these will be dropped from the data frame; only 12 observations are dropped in total.

TABLE 3. Data reference contains index and count.

S. No	Index	Count
14	position	1
15	plan_configuration	1
28	damage_grade	12
29	technical_solution_proposed	12

The following section explain about the Exploratory Analysis between the Damage Grade (Target feature) and features (Categorical and Numerical variables) in Earthquake Damage Dataset.

C. EXPLORATORY ANALYSIS

To start, we look at the distribution of the target variable, and we can see the occurrences of each grade increases with the classification. Such that Grade 5 occurs the most frequently in the Dataset while Grade 1 appears the least frequently. As this is a classification problem, we can see that it is unbalanced, with different grades accounting for very different proportions of observations. The below Figure -2 shows that the Distribution of Damage Grade.

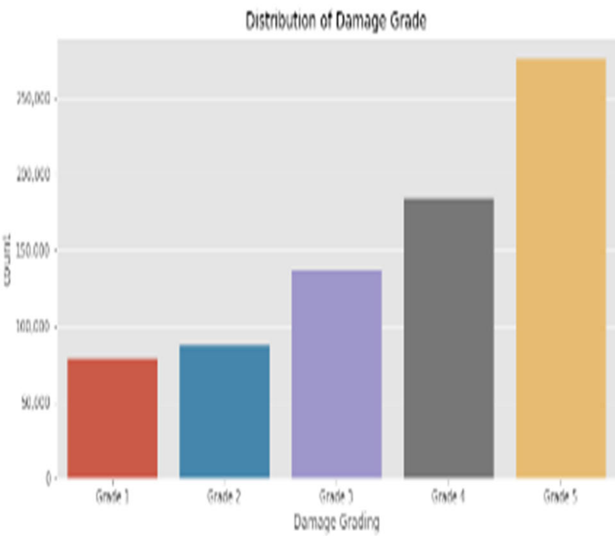


FIGURE 2. Distribution of damage grade.

D. CATEGORICAL VARIABLES

The output below shows the non-numeric fields in Dataset. These variables will be explored in turn to understand how they relate to the damage grade classification. The first variable to analyze is the district; from the plot below the most districts are spread relatively evenly across all damage grades. There are a few examples of districts that are mostly associated with grade 5. The Figure-3 shows that Distribution of Damage Grade by District.

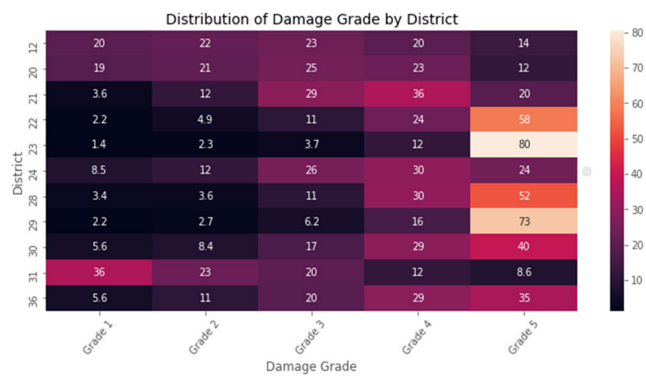


FIGURE 3. Distribution of damage grade by district.

The plot below Figure-4 shows that the association between Land Surface Type condition and damage grade.

The next categorical variable is "Foundation type". There are five entries in the foundation type, Mud mortar-Stone/Brick, Bamboo/Timber, Cement-Stone/Brick,

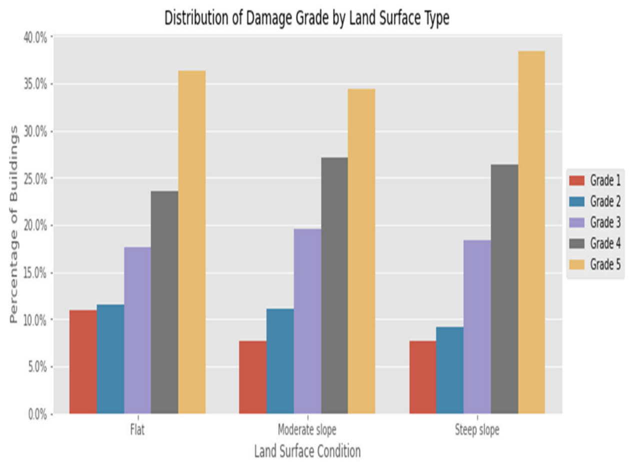


FIGURE 4. **Distribution of damage grade by land surface type.**

RC (Reinforced Concrete) and other. The below Table-4 shows that different types of Foundation Types among all the 5th option being a grouping of less popular options. The below Figure-5 shows the Distribution of Damage Grade by Foundation Type and Visually this variable appears to have some degree of predictive power.

TABLE 4. **Different types of foundation type.**

Mud mortar-Stone/Brick	628705
Bamboo/Timber	57472
Cement-Stone/Brick	39245
RC	32120
Other	4552

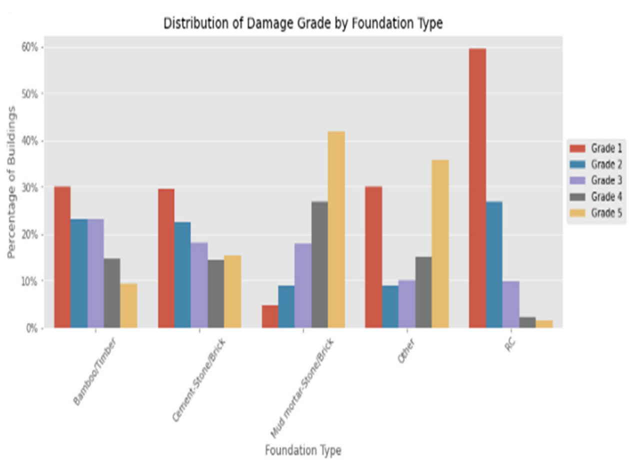


FIGURE 5. **Distribution of damage grade by foundation type.**

According to the ''roof type variable'', there are three different types of roofs, Bamboo-Timber/ light roof, Bamboo-Timber/ heavy roof, RCC (reinforced cement concrete)/RB (Reinforced Brick)/RBC(Reinforced brick concrete). The vast majority of buildings have the Bamboo-Timber/ light roof, but there are still over 200k buildings with timber heavy and just over 5k with RCC. The spread of building between roof type and Damage Grade is quite similar for the types with Bamboo and timber. Roof type RCC actually is most commonly associated with grade 1 and then 2. It is minimally associated with grades 4 and 5. The below Figure-6 shows that Distribution of Damage Grade by roof type.

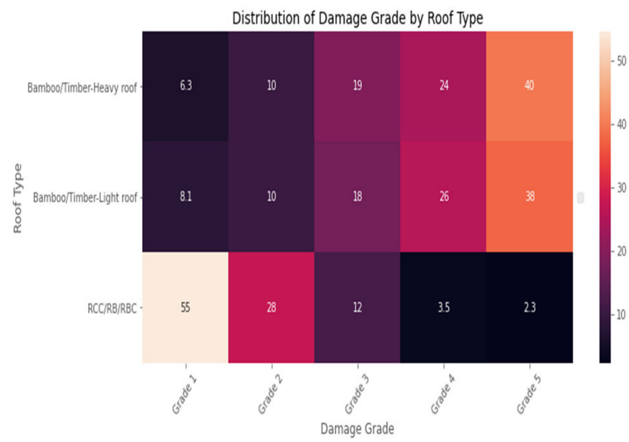


FIGURE 6. **Distribution of damage grade by roof type.**

There are multiple types of ground floor material used across the buildings, one the most popular is Mud, which can be seen below. RC and Brick are distant second/ third place but occur quite frequently. The below Table-5 indicates that the different types of ground floor material.

TABLE 5. **Types of ground floor material.**

Mud	618205
RC	73149
Brick/Stone	66093
Timber	3594
Other	1053

The below Figure-7 shows distribution of damage by Ground Floor Type.

The two outputs (Table-6 (Distribution of Damage % Over the Other Floor Type) & Figure-8(Distribution of Damage Grade by Other Floor Type)) below are focused on the 'other_floor_type' variable; see that timber bamboo mud is a very common flooring material, existing in over 63% of homes. There is also an option called not applicable, indicating that some buildings only have a ground floor, which might be worth creating a feature to extract.

The next variable is ''Position'', the instances of this variable is Not attached, Attached-1 side, Attached-2 side and Attached-3 side. The association between position and Damage Grade of buildings differs less when compared with No attachment and buildings with an attachment on one side. The below Figure-9 shows that the Distribution of Damage Grade

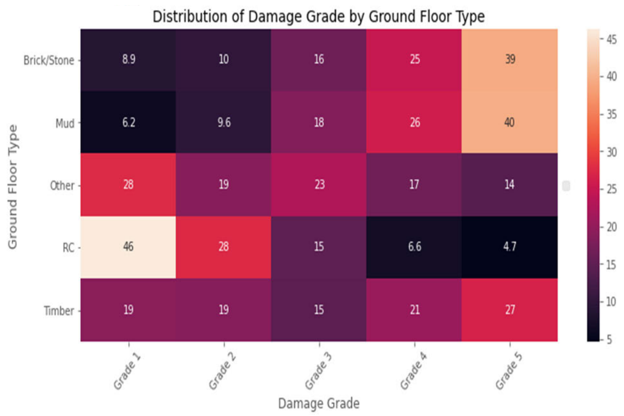


FIGURE 7. Distribution of damage grade by ground floor type.

TABLE 6. Distribution of damage % over other floor type.

Timber/Bamboo-Mud	0.638907
Timber-Planck	0.162216
Not applicable	0.155914
RCC/RB/RBC	0.042963

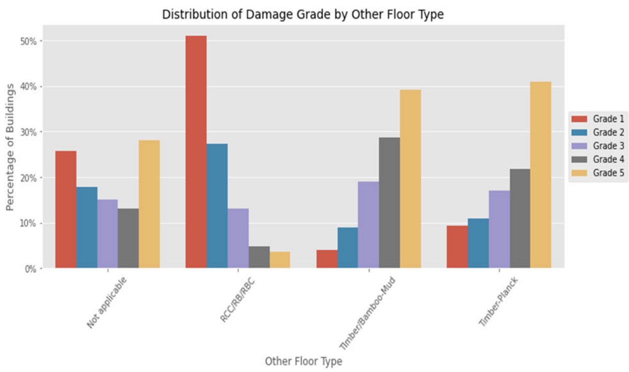


FIGURE 8. Distribution of damage grade by other floor type.

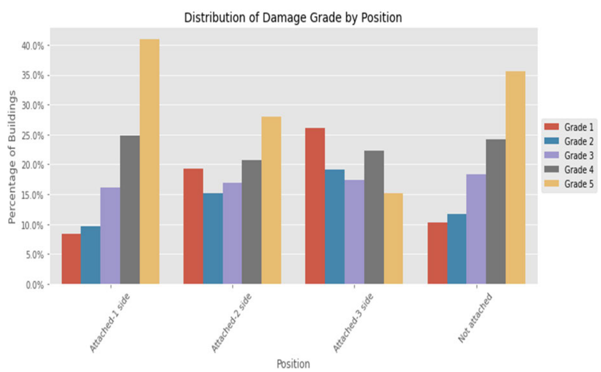


FIGURE 9. Distribution of damage grade by position.

by Position. Table-7 shows that the Comparison between with 2 and 3 side attachments and without attachments.

TABLE 7. Comparsion between non-attachment and with attachment.

Not attached	0.793134
Attached-1 side	0.169836
Attached-2 side	0.035311
Attached-3 side	0.001719

The plan configuration represents the building's general shape, and the below Table-8 shows that the ten different types of plot shapes. The most popular type of plot shape is rectangular, followed by square and then L-shape gives that in the Figure-10.

TABLE 8. The ten different types of plot shapes.

Rectangular	731246
Square	17576
L-shape	10079
T-shape	969
Multi-projected	940
Others	518
U-shape	448
E-shape	140
Building with Central Courtyard	98
H-shape	80

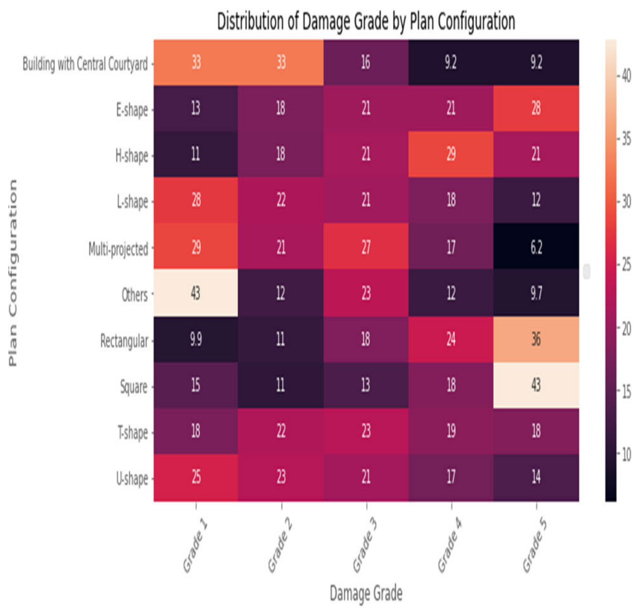


FIGURE 10. Distribution of damage grade by plan configuration.

The condition field appears to evaluate the standing of the building after the Earthquake. Unfortunately, there are several common recordings, all pointing towards some form of damage (only 8% were not damaged). Now the below Table-9 illustrations that the Damage Parameters Range and the Damage Repair Range in Table-10.

TABLE 9. Damage parameters range.

Damaged-Not used	207968
Damaged-Rubble unclear	125650
Damaged-Used in risk	123843
Damaged-Repaired and used	107791
Damaged-Rubble clear	102191
Not damaged	61139
Damaged-Rubble Clear-New building built	33130
Covered by landslide	382

TABLE 10. Damage repair range.

Reconstruction	470219
Major repair	129415
Minor repair	110605
No need	51855

The final categorical variable identifies the scale of damage caused to the building into four groups going from no need to intervene to complete reconstruction. The majority of buildings required reconstruction or some degree of repair. The plot shows a nice pattern, ultimately Grade 1 buildings not requiring any intervention, grades 2 and 3 requiring minor and major repair respectively, and then grades four and 5 requiring reconstruction. The Table-11 shows that the all parameters statistical ranges. The below Figure-11 shows that the Distribution of Damage Grade by Condition.

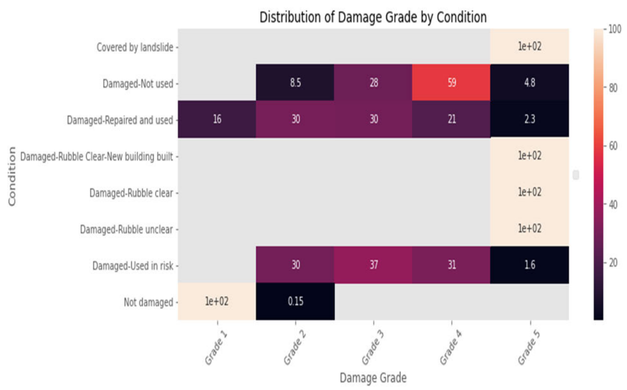


FIGURE 11. Distribution of damage grade by condition.

E. NUMERICAL VARIABLES

In this Section, exploring the relationship between damage grade and the numerical variables. From the dataset the following are the Numerical variables: variables, which have the 'has' prefix, Building age, Number of rooms and Height of Building. The output below shows the descriptive statistics about the numerical variables in our Dataset. The first thing to notice is that all the variables, which have the 'has' prefix, are binary variables. Building age has an 8-year gap between the mean and median, and the maximum value that the age is recorded as 999 years old.

The plot below looks at the number of rooms a building had before and after the Earthquake. Focusing on the left plot (before Earthquake) it shows the distribution of rooms that building had. Focusing on the right plot (after Earthquake); we can see a pattern that higher dam age grades associate to lower distribution of rooms than befo re. (Figure-12). In Figure-13 shows that the ECDF (These graphs require continuous variables and allow you to derive percentiles and other distribution properties. This function is also known as the empirical CDF or ECDF. If you measure the same characteristic in multiple samples, you can use empirical CDF plots to compare the sample distributions. Empirical Cumulative Distribution Function) for Building age.

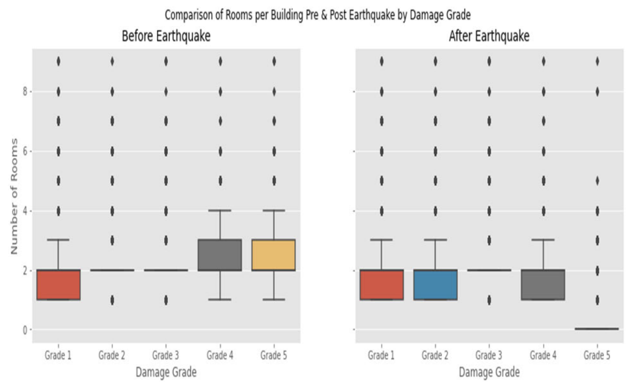


FIGURE 12. The plots of empirical cumulative distribution for the Building's age split by damage grade.

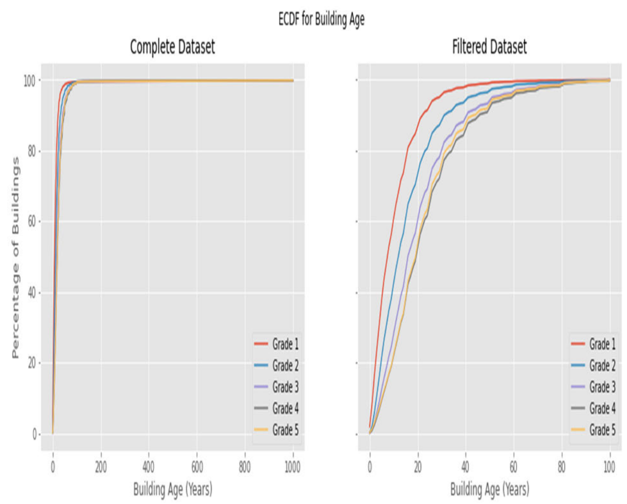


FIGURE 13. ECDF for building age.

The plot below (Figure-14) shows the distribution of the building's plinth area by grade. Ignoring the height of the bars between grades, we can see that the building area generally, has the same distribution with a peak of 250 sq. ft.

The plot below (Figure-15) shows the distribution of building heights before and after the Earthquake by damage grade.

TABLE 11. All parameters statistica ranges.

Parameters	Mean	Std	Min	25%	50%	75%	Max
count_floors_pre_eq	2.087787	0.6551	1	2	2	2	9
count_floors_post_eq	1.25205	1.06328	0	0	1	2	9
age_building	24.325031	65.0346	0	9	16	27	999
plinth_area_sq_ft	406.67367	226.78	70	280	358	470	5000
height_ft_pre_eq	16.049424	5.4939	6	12	16	18	99
height_ft_post_eq	9.868785	8.57422	0	0	11	16	99
has_superstructure_adobe_mud	0.042402	0.2015	0	0	0	0	1
has_superstructure_mud_mortar_stone	0.800269	0.3998	0	1	1	1	1
has_superstructure_stone_flag	0.035122	0.18409	0	0	0	0	1
has_superstructure_cement_mortar_stone	0.015816	0.12476	0	0	0	0	1
has_superstructure_mud_mortar_brick	0.022962	0.14978	0	0	0	0	1
has_superstructure_cement_mortar_brick	0.071527	0.2577	0	0	0	0	1
has_superstructure_timber	0.25877	0.43796	0	0	0	1	1
has_superstructure_bamboo	0.080484	0.27204	0	0	0	0	1
has_superstructure_rc_non_engineered	0.039794	0.19548	0	0	0	0	1
has_superstructure_rc_engineered	0.016386	0.12696	0	0	0	0	1
has_superstructure_other	0.012026	0.109	0	0	0	0	1

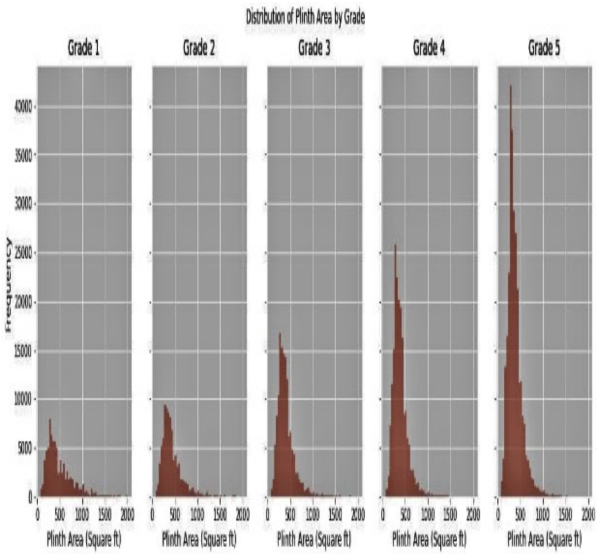


FIGURE 14. Distribution of Plinth Area by Grade.

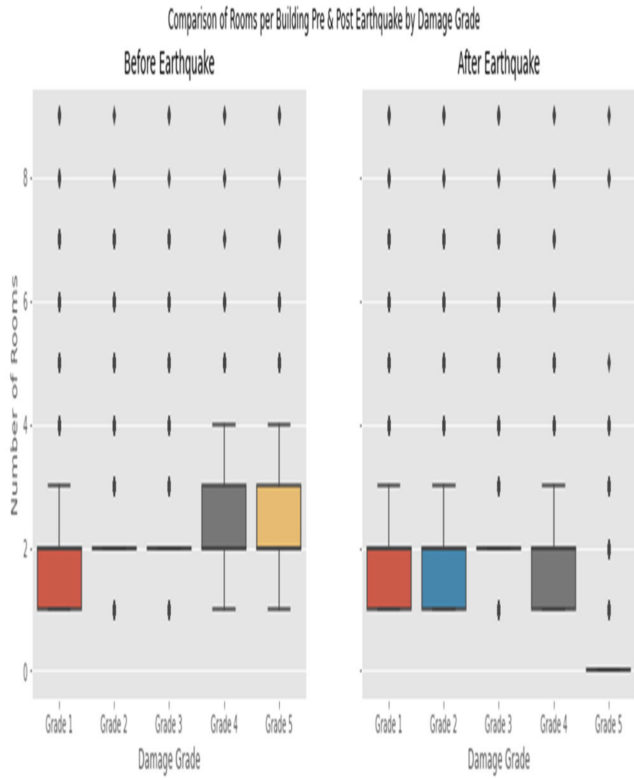


FIGURE 15. Comparison of rooms per building pre & post earthquake by damage grade.

For damage grades 1-3, the general distribution building heights are unchanged between before and after the earthquake Damage grade 4 displays a slight reduction in height between the two time periods, while damage grade 5 drops to zero for the bulk of the distribution, indicating that these buildings completely fell down.

The below plot shows (Figure-16) that summarization of the occurrence of each of the superstructure variables.

In this next section, explain about the proposed model called Hybrid Future Selection Methodology (HFSM) for finding the abstract features. HFSM contains two methods: Mutual Information (MI) and Rough Set Theory (RST).

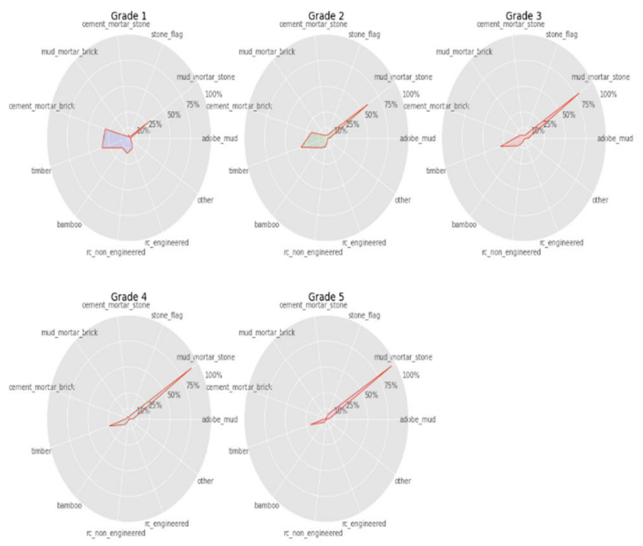


FIGURE 16. The plot summarizes of the mean occurrence of each of the superstructure variables.

IV. BUILT IN METHODOLOGY

In this section, explain about Built-in methods for feature subset selection, classification and predicting strength status of the buildings.

A. FEATURE ENGINEERING

Before applying the model on the dataset, the most important step is to examining the data. This increases the efficiency of the model. Basically, the examining of data can be done by using any pre-processing step. Pre-process the data gives the better results and increases the efficiency of the model. The following are the procedure for pre-processing. There are two files, which contain the features of different properties of the problem, but these are used for problem solving or training a model. So, merge data with the train data, which helps us to build a robust model (Data Merging). The data is not balanced, which affects the model's accuracy. So "SMOTE" algorithm used to make the data balanced in all the classes. Also, convert all the non-numeric columns into numeric columns (Data balancing). The missing values in the Dataset create confusion and affect the results. Therefore, as per the intuition, those values are filled using zero (Missing values). If there are two or more columns that are highly correlated with each other, then it really affects the model's accuracy. So, removed highly correlated columns from the data (High-Correlated). If there are duplicates in the training data, then there is a high chance of overfitting, which affects the accuracy of the model or method when tested on new data. So, removed duplicates from the data (Duplicate Attribute). If there are any constant columns in the Dataset, there is no use to use that feature to train the data (Constant values).

Now, in this paper, we proposed a Hybrid Future Selection Methodology (HFSM) for finding the abstract features. HFSM contains two methods: Mutual Information (MI) and Rough Set Theory (RST). MI gives how much a parameter is correlated to the other parameters in the Dataset. The RS pattern helps us to find the approximate value between the screen parameters.

Algorithm 1 Pre-Processing (Mutual Information)

INPUT:

$$\text{Data Set } ID_i = \sum_{i=0}^{n} A_i$$

Where

A is the instances of the attributes and i = 0, 1, 2, 3.....n.

n is number of elements in the Dataset ID_i.

OUTPUT:

Data set with most Co-related and ranks with in the threshold value and Data set with Low approximate value.

$$OD_j = \sum_{j=0}^{n-m} A_j$$

where **m** is the number belongs to how many elements are reduced from the ID_i.

STEP-1:

$$\text{Take } ID_i = \sum_{i=0}^{n} A_i \text{as input}$$

STEP-2:

Find the most co-related among the attributes using the following formula.

$$MI(p,q) \sum_{p=n} \sum_{q=n} pb(p,q) \log \frac{pb(p,q)}{pb\,(p)\,pb(q)}$$

Here,

pb (p, q): Represents the Probability of p over q.

pb (p): Represents the Probability of p.

pb (q): Represents the Probability of q.

STEP-3:

For each attribute in the ID_i

Assign Ranks(R) based on Correlated values obtain in step-2.

STEP-4:

For each attribute in the ID_i

Fix the Threshold values (TV).

STEP-5:

For each attribute in the ID_i

If R<TV then

$$\text{Send to } OD_j = \sum_{j=0}^{n-m} A_j$$

Else

$$\text{Send to remove set } \mathrm{RM}_t = \sum_{t=0}^{m} R_t$$

STEP-6: stop.

Algorithm 2 Pre-Processing (Rough Set Theory)

STEP-1:

Take $ID_i = \sum_{i=0}^{n} A_i$ as input with n attributes,

Where

'A' is the attribute name and 'i' is the index or instances.

STEP- 2:

The Decision System (DSM) defined as follows.

$$DSM : OD_j = (\mathrm{DI}, \mathrm{DI}_c \cup \{\mathrm{D}_a\})$$

Where,

DI_c is a non-empty finite set of attributes,

$$\forall a \in DI_c, a : \mathrm{DI} \rightarrow V_a$$

The value set of A is Decision attributes.

The elements are called DI_c conditional attributes and OD_j are the Final attribute Set.

STEP-3:

For each attribute in ID_i repeat **the step A.**

STEP-A:

Estimate the dependency degree change (D).

STEP-4:

For each attribute in the ID_i

Fix the Threshold values (TV).

STEP-5:

For each attribute in the ID_i

If D<TV then

$$\text{Send to } OD_j = \sum_{j=0}^{n-m} A_j$$

Else

$$\text{Send to remove set } \mathrm{RM}_t = \sum_{t=0}^{m} R_t$$

Here **m** is the number belongs to how many elements are reduced from the ID_i.

STEP-6: STOP.

Algorithm-1 about Mutual Information and Algorithm-2 about Rough Set Theory. These two are defined the procedure for identifying the most significant attributes from Dataset. Now the below figure-17 shows that the overall procedure of the Hybrid Future Selection Methodology (HFSM).

Now apply the Pre-processing algorithms-1&2 on a dataset (Gorkha earthquake in April 2015), to get the two attributes called Net Rooms and Net Height (Both are most significate attributes for the built model). Both are observed before and after Earthquake and used for assessing the damage levels (LE). Net Rooms (Number of rooms after the earthquake - Number of rooms before the earthquake) and Net Height (Building height after the earthquake - Building height before the earthquake) are two features used for predicting the Damage Grade.

Next, explain about Built in Methods for classification and predicting strength status of the buildings. SVM methodology is used for classification and ANN is used for predicting status of damage levels.

B. CLASSIFICATION AND PREDICATING ENGINEERING

1) SVM: SUPPORT VECTOR MACHINE

In 1995 Cortes et al. [37] introduced one of the best- supervised optimization classifier called support vector machine (SVM).The working of SVM is based on statistical learning theory and used for both classification and Regression experiments. SVM create the Hyper-plane and used for identifying the correct classification (two classes). From the dataset the two classes are Damaged and Undamaged building structures. However, the problem is "How can we identify the right hyper-plane?" The solution for the problem is called Margin, the gap between nearest items and plane. The Margin is equivalent to the distance. If we select a hyper-plane having a low margin, then there is a high chance of miss-classification otherwise perfect classification. SVM classification is robust to outliers. Suppose the plot hyperplane is not a linear hyper-plane between the two-class and SVM using the Kernel trick technique. That means a kernel function proceeds truncated dimensional input-space & makes it over to a higher-dimensional space. The main aim is an innovation of hyperplane in M-dimensional space where M is the number of structures. Using hyperplane, the given data points are classified as true-class and false-class in the incomparable form. For analysis, the given problem needs ideal hyperplane.

This plane suggested that this is a good classifier, defined as:

$$\begin{aligned} &\text{Minimize to : } \frac{1}{2}\|v\|^2 \\ &\text{Subject to : } y_i\left(\mathrm{V}^T.X_i + b\right) \geq 1, i = 1, 2, \ldots\ldots.\mathrm{m} \\ &\qquad y_i \in \{1, -1\}, \mathrm{X}_i \in R^m \end{aligned} \tag{1}$$

From the above formula, "v" characterized as a vector with "m-dimensional space" and assign a name to each data points as $\mathrm{X_i}$ used for classification of "m" elasticities count, scalar [39].

2) ANN: ARTIFICIAL NEURAL NETWORKS

ANN is a biological methodology, prearranged in a Hierarchical (layer by layer) format. ANN is used for solving pattern recognition problems and is highly involved in estimating the status of damage in building structures based on modal considerations like natural frequencies, damping ratio, and any human or environmental errors. Ghazali et al. [40] and Chu et al. [41], [42], [43] define the two transfer functions: Input function and Activation functions. These functions are worked on the given input signals $P_1, \mathrm{P}_2, \mathrm{P}_3, \ldots\ldots\ldots., \mathrm{P}_n$. The input function first processes the input outcome Net_j(y)

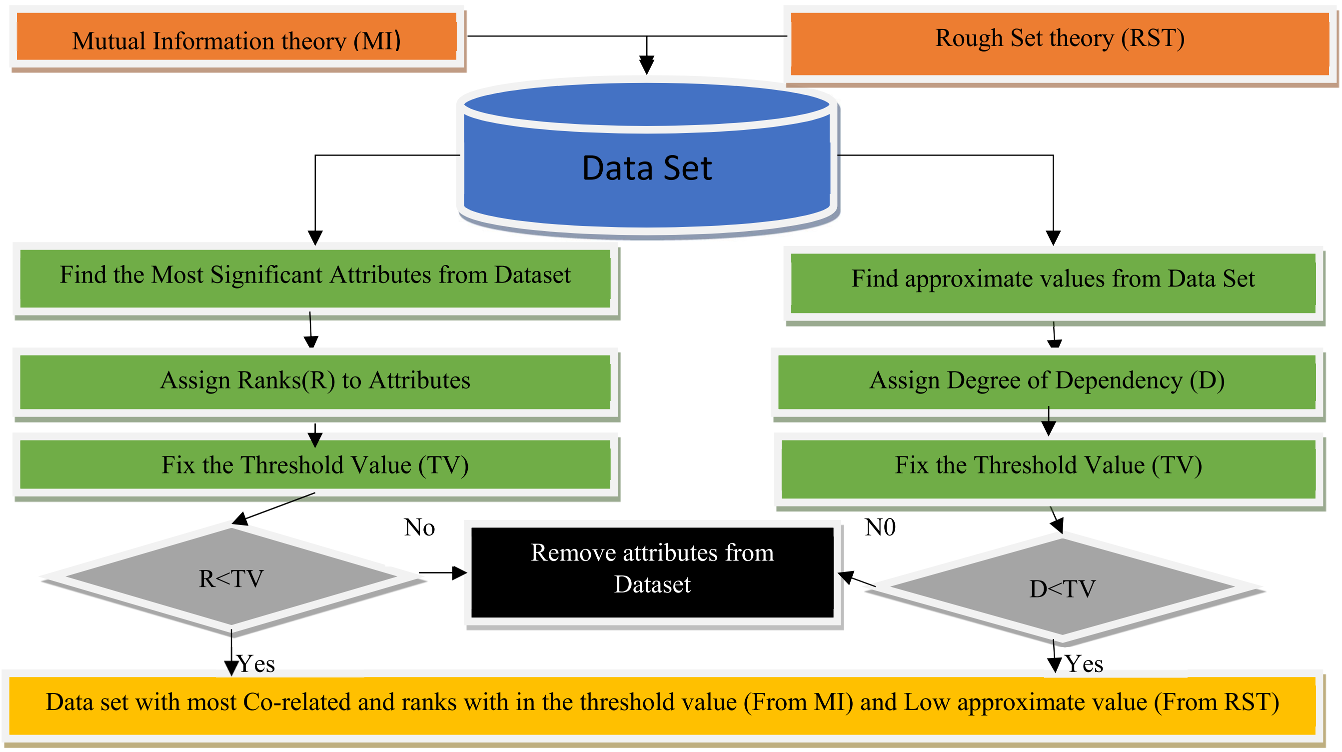


FIGURE 17. The overall procedure of the hybrid future selection methodology (HFSM).

as an activation function and producing the output signal. Wasserman et al. [44] say that the learning mechanisms for self-learning and multi-layer networks are categorized as supervised and unsupervised learning. Feedforward is the first stage of operation of the Feed Forward-Back Propagation NN (FF-BPNN).

The operation determines the neuron is output in the Hidden layer $HD_n(y)$ for a given P(y) input vector and y^{th} pattern of the training.

$$HD_n(y) = f\left[Net_h(y)\right] \tag{2}$$

$$\text{Now } Net_h \text{ is defined as } Net_h = \sum W_{hi}(y).P_i(y) \tag{3}$$

The below Figure-18 shows that the 3-layer BPNN. In the same way, the Yield element OL in the output layer $LO_o(y)$ is $LO_o(y) = f[Net_o(y)]$.

Now $Net_o(y)$ is defined as

$$Net_o(y) = \sum W_{oh}.f\left[\sum W_{hi}(y).P_i(y)\right] \tag{4}$$

"Error Back Propagation" is the next stage of the BPNN. Now, $EF(o)$ is the Error Function and defined as

$$EF(o) = \frac{1}{2}\sum [T_o(y) - OL_o(y)]^2 \tag{5}$$

Some external forces, changes in temperature and changes in load Kao et al. [45] observed the changes in civil structures, developed the method contains the two-steps. one is the identification of the System and second one is used to predicate Structural Damage Detection based on vibration response. Sahin et al. [46] plan about the hybrid scheme for finding the place and severity of damages in buildings.

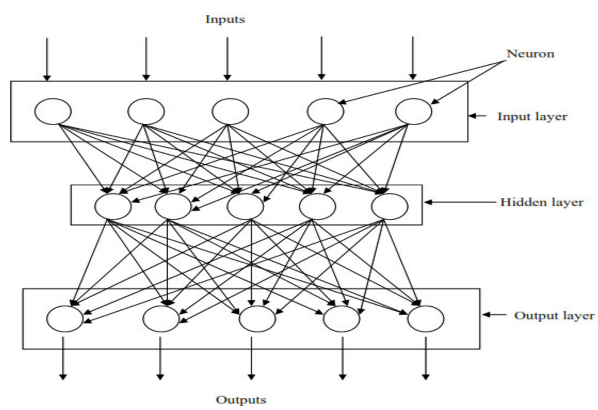


FIGURE 18. The 3-layer BPNN.

For analysis of the model,uses the vibration data (Strain Device), changes in model properties and Natural Frequencies (Accelerometer) as a parameters. Xu et al. [47], Back-Propagation ANN(BPANN) for calculate approximately Damping Coefficient, Stiffness deprived of Eigen Value and it's analysis done based on the 5-story structure. Lee et al. [48], ANN used to estimate the severity, location of the damage. This method works on the bases of changes in the parameters of structure and analysis done on civil bridges. Ahadzadeh et al. [49], ANN with Co-occurrence Matrix used to estimate the changes in the building strucutres with the accuracy of 62% in case of building experienced the Earthquake.

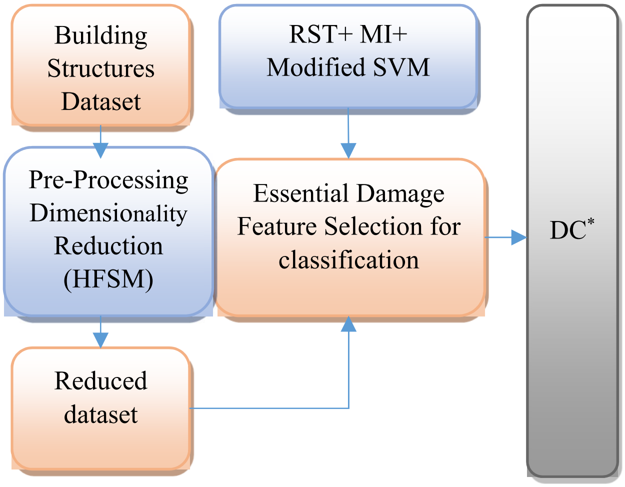


FIGURE 19. Damage classification (unsupervised) method.

The above Figure-19 shows that the methodlogy for building Damage Classification. In the next section, explain about the Hybrid Machine Learning Technique (HMLT) for assess the damage levels in the building. Inside this, a pre-processing pipeline will be set up to prepare the data for training. A test dataset will be extracted from the main data, to give us the opportunity to assess how the models perform on completely new data. The models will be evaluated on the test data using the evaluation metric F1.

V. PROPOSED HYBRID MACHINE LEARNING TECHNIQUE (HMLT)

The below Figure-20 shows that the hybrid method called as Hybrid Machine Learning Technique (HMLT). The methodology is first collecting the Building structures data, that contains both damage and undamaged information. Earthquake signals are used for creating the dataset. From this dataset the major task is identifying the most significant features using one hybrid Pre-Processing method called HFSM.

The outcome of this methodology is most significant features for getting the good result with low power. Now apply classification method on these features for finding the hyper plane, that are separated the features into two sets (damage and undamaged) then apply any predicting the methods called Hybrid Machine Learning Technique (HMLT) for assess the damage levels in the building (Figure-20).

The workflow of the HMLT, first Start, next find the most co-related features using the Pre-processing method (HFSM). These features contain the data for damage, features of damage in related to buildings and calculating similarity in Damage feature gives the damage type, building types, disaster data. Then combined all features and calculate the weights of the damage feature attributes using any classification, prediction methods for estimate the damage levels.

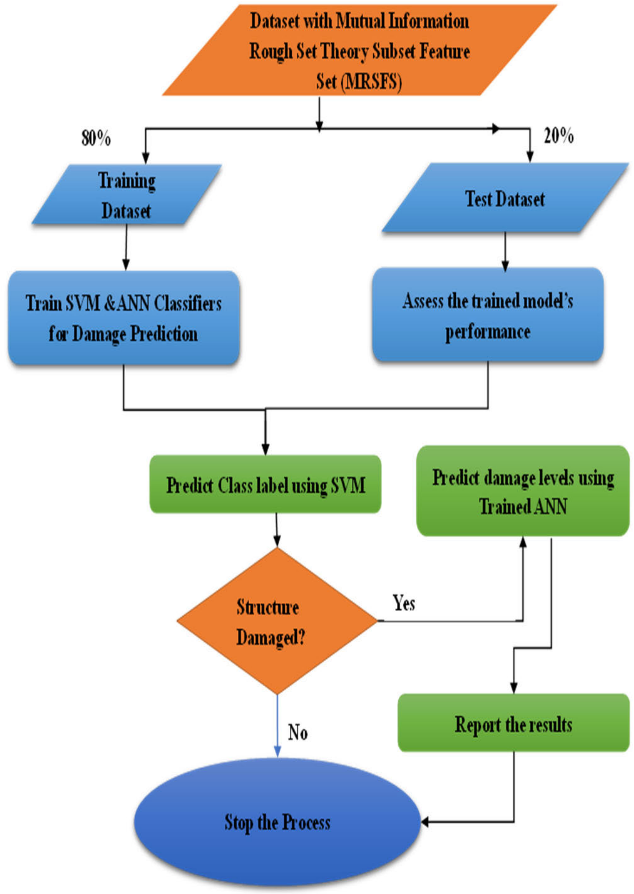


FIGURE 20. Hybrid machine learning technique (HMLT).

Now in the next section, explain about the results of the above ALGORITHM-3 and the detail explanation about the comparison between proposed and classifiers. This comparison says that our proposed method gives better accuracy. The models will be evaluated on the test data using the evaluation metric F1.

VI. RESULTS

A. TRAINING AND TEST DATASET

A test set is created to help evaluate the performance of predictive models, which will be developed. It provides the opportunity to assess how the model performs on unseen data. The test set will account for 20% of the observations, chosen at random but stratified around the target variable to ensure that the proportions are the same in the training and testing data. The Table-12 shows that this has been achieved and cannot see that the training and test data frames have very similar proportions of observations for each damage grade

B. PRE-PROCESSING PIPELINE

In this section, focused on preparing the data for the proposed methodology. The outcome of this step is identifying the most significant attributes, which are used for getting better results. The following rules will be applied on Dataset, Label Encode

Algorithm 3 Hybrid Machine Learning Technique (HMLT)

INPUT:

$$OD_j = \sum_{j=0}^{n-m} A_j$$

Where '**m**' is the number Reduced Features and total number of Features ID_i are 'n'.

OUTPUT:

Predicate the Damage grades in ID_i.

STEP-1:

for each j=0ton-m

Find Hyperplane with m-dimensional

Minimize to: $\frac{1}{2}\|v\|^2$

Subject to : $y_i\left(V^T.X_i + b\right) \geq 1, i = 1, 2, \ldots\ldots.m$

$y_i \in \{1, -1\}, X_i \in R^m$

STEP-2:

for each j=0 to n-m

Draw Margin.

STEP-3:

If (Hyperplane== low Margin)

Miss-classification

Else

Perfect classification.

STEP-4:

Take and move with classification (Damage building data) to next step.

STEP-5:

Using FF-BPNN

Find the damage grades.

STEP-6:

For each neuron

Find $HD_n(y) = f\left[Net_h(y)\right]$

P(y) Input vector and y^{th} pattern.

Now Net_h is defined $Net_h = \sum W_{hi}(y).P_i(y)$

STEP-7:

For each neuron

find $LO_o(y) = f[Net_o(y)]$.

$$Net_o(y) = \sum W_{oh}.f[\sum W_{hi}(y).P_i(y)]$$

.

STEP-8:

$$\text{Estimate } EF(o) = \frac{1}{2}\sum [T_o(y) - OL_o(y)]^2$$

STEP-9:

Predicate the Damage grades in ID_i.

STEP-10:

STOP.

Target Variable, ensure variables have the correct data types, Covert nominal variables to numeric with one hot encoding,

TABLE 12. Training and testing dataset.

S.NO	Index	Train	Test
0	Grade 5	0.361853	0.36185
1	Grade 4	0.241235	0.24124
2	Grade 3	0.178995	0.179
3	Grade 2	0.114497	0.11449
4	Grade 1	0.103419	0.10342

restrain outliers, Centre and scale all numerical variables, Remove variables with either no or minimal variance. Firstly, we encode the objects y_train and y_test to be value labels ranging from zero to four, corresponding to Grades 1 - 5.

The second step is to create a pipeline of processing steps, which can be applied to the predictor features. These are all combined into a single pipe, which easily allows the same processing step to be applied to different data frame. The print out below shows that with have significantly increased the dimensionality of our Dataset as a result of one hot encoding the categorical variables. Figure-21 shows that the anomaly detection results. Table-13 below shows that the various classifiers and its performances in training followed by ten-fold and five-fold. Based on the results, used fivefold cross validation only.

TABLE 13. Classifiers vs the performances in tenfold and fivefold crass validation.

DS	METHODS							
	SVM				ANN			
	T	TCV	T	FCV	T	TCV	T	TCV
	98.43	90.62	99.02	91.02	96.8	89.06	100	90.67

Where T: Training, TCV: Tenfold cross validation, FCV: Fivefold cross validation, DS: Data Set.

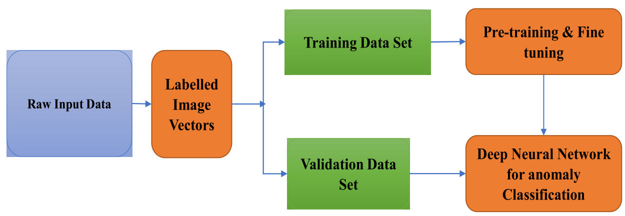


FIGURE 21. Anomaly detection results.

After getting an understanding of the dataset, we used our novel Feature Selection algorithm to select relevant features to reduce processing time and get better results with our models. The below Table-14 shows that the pre-processing Dataset.

C. FIVEFOLD CROSS VALIDATION

Sample size, number of features, data distribution and number of classes plays an important role for the optimal number of folds. The p-fold cross validation assessed as unbiased

TABLE 14. Pre-processing dataset.

Before pre-processing	After pre-processing
Before pre-processing, there were 609675 rows and 31 columns.	After pre-processing, there are 609675 rows and 69 columns.

containing expected value of error arising from design of sample of size. Two fold cross validation are unbiased estimate of error for classifier trained by half of given N data samples and leave one by N-1 samples. If reduce the number of folds, a negative bias appears, because use a smaller training set. If the number of samples is high enough, all cross validation estimates should give similar values. If the size of the fold is high then the variance is also high. According to Efron et al. [50], variance dominates the samples. Therefore, less variance is good chose for getting better results. Krizek et al. [51], recurrent two-fold cross validation was testified to prime to the best permanence and presentation of wrapper methods. Bailey et al. [52] suggested that the low variance and higher bias than leave-one-out cross validation. So here, considered the 5 fold cross validation only. Fivefold cross validation will be used as the strategy, which will provide confidence that we are not overfitting and allow us to make fair comparisons between competing models. Ultimately, this approach splits the data into 5 random partitions and builds 5 models, for each model it uses 4 partitions of data to train the models and the fifth for testing.

D. COMPARISONS

In this section, explain about the metrics and comparison of metrics with proposed classifiers.

TABLE 15. F1-score of classifiers.

Classifier	F1-score
KNN: k-nearest neighbors	0.79
GBM: Gradient Boosted Machines.	0.89
RF: Random Forest	0.88
Logistic Regression(LR)	0.88
SGD Classification	0.90
Hybrid Machine Learning Technique (HMLT)	0.91

1) DEFINITION OF METRIC

In the present paper, we used a combination of MI and RST for Feature Selection and subsequently used SVM to classify if a building is damaged, followed by employing ANN to assess the severity or level of the damage.

In order to evaluate and understand the results obtained, we are using four performance metrics:

- Accuracy
- Precision
- Recall
- F1 Score

Now, for any prediction the model is making, it is one of the four:

- A True Positive (TP): It is positive, and the model correctly classifies it as positive
- A True Negative (TN): It is negative, and the model correctly classifies it as negative
- A False Positive (FP): It is negative, but the model wrongly classifies it as positive
- A False Negative (FN): It is positive, but the model wrongly classifies it as negative.

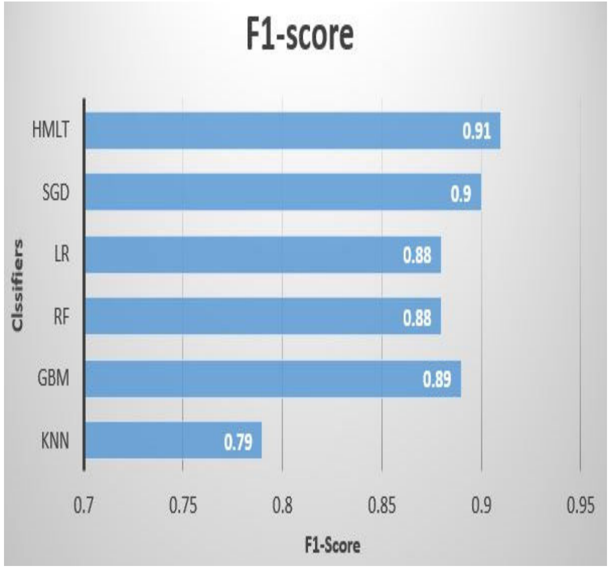


FIGURE 22. F1-Score value of proposed HMLT with other classifiers.

TABLE 16. Comparison between the different classification and its accuracy.

Classifier	Accuracy
RCHSVM	0.76
KNN: k-nearest neighbors	0.79
SVM	0.85
ANN	0.86
GBM: Gradient Boosted Machines.	0.88
RF: Random Forest	0.89
Logistic Regression	0.89
SMO	0.90
LIBSVM	0.90
SGD Classification	0.91
Hybrid Machine Learning Technique (HMLT)	0.92

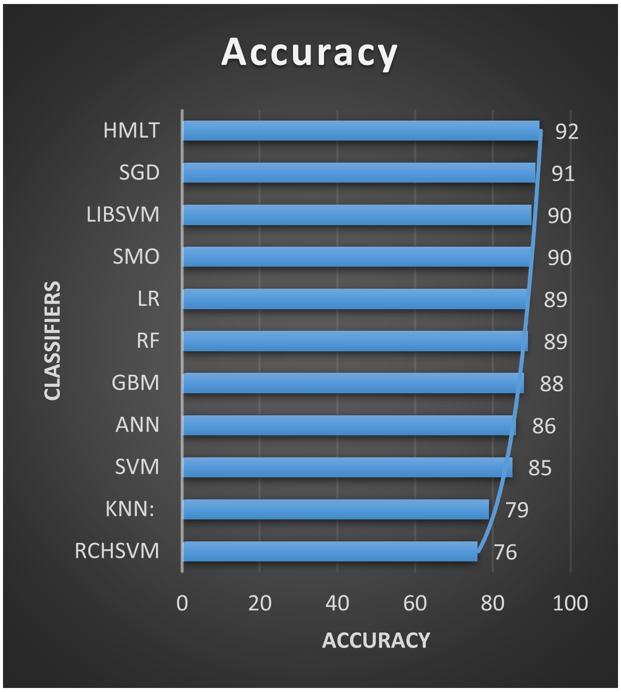


FIGURE 23. **Accuracy of proposed HMLT with other classifiers.**

Accuracy:

Accuracy informally is defined as the percentage of values the model got correct.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision:

Precision is the percentage of correctly labelled positives out of all positively labelled points.

$$Precision = \frac{TP}{TP + FP}$$

Recall:

Also known as sensitivity, it measures the percentage of correctly labelled positives out of all the actual positives in the dataset.

$$Recall = \frac{TP}{TP + FN}$$

F1 Score:

F1 Score is the harmonic average of Precision and Recall.

$$F1Score = \frac{2 * (Precision) * (Recall)}{Precision + Recall}$$

Tables-15&16 shows the F1 Score and accuracy obtained using Classifiers and compares these with the proposed method's accuracy.

2) CLASSIFIERS VS F1 SCORE VALUES

The models are built across the same five folds of data, and the output below shows how each model performs using the F1 score evaluation metric. This present the summarized evaluation, indicating how the model generally performed across all of the folds. The below Table-15 shows that the F1-Score of Classifiers. The following are the some of the observations, KNN model: It performed the worst, achieving an F1 score of 79.2%; the low standard deviation score indicates that similar performance was achieved across the five folds. SGD Classification model: This model achieved an F1 score of 90%, just over a 10% point increase relative to the simple KNN model. The standard deviation score for this model also indicates similar performance across the folds. Gradient Boosted Machine model: This model achieves an F1 score of 88.3%, which underperforms against the random forest but performs strongly against the KNN model. The proposed Hybrid Machine Learning Technique (HMLT) model achieves an F1 score of 91%, which is higher than other existing classifier.

The below Figure-22 shows that the F1-Score comparison between existing classifiers and proposed method (HMLT).

3) CLASSIFIERS VS ACCURACY

Now the below Table-16 shows that the comparison of accuracy between Classifiers and proposed Classifier. The results analysis, says that our proposed HMLT method having high accuracy then others. That means our proposed method got better results. The Figure-23 shows that the Accuracy of Proposed Classifier with other Classifiers.

VII. CONCLUSION AND FUTURE WORK

Structure Health Monitoring (SHM) is one of the automation system for forecasting or monitor the strength status of buildings. Several researchers are focusing on this field and develop hybrid methods for Monitor the strength status of buildings. This manuscript presents an interesting idea for predicting applications to monitor the strength status of buildings. Now in this article, develop Hybrid Machine Learning Technique (HMLT) for monitor the strength status of buildings. The combined features of Mutual information (MI) and Rough Set Theory (RST) are used for feature selection and the optimized classifiers such as Support Vector Machine (SVM) and Artificial Neural Networks (ANN) are used for classification, predicting the status of the building strength. Apply various classifiers in training dataset (80%) with tenfold and fivefold. But the results says that the fivefold cross validation In order to evaluate and understand the results obtained, we are using four performance metrics: Accuracy, Precision, Recall, and F1 Score with fivefold cross validation on damage dataset and also F1-Score values of the Classifiers are compared with proposed Classifier. The following are the some of the observations, KNN model achieving an F1 score of 79.2%,, SGD Classification model achieved an F1 score of 90%, Gradient Boosted Machine model achieves an F1 score of 88.3%, which underperforms against the random forest but performs strongly against the KNN model. The proposed Hybrid Machine Learning Technique (HMLT) model achieves an F1 score of 91%, which is higher. The results analysis, says that our proposed HMLT method having high accuracy. i.e., our proposed method got better results. According to these comparisons, the suggested HMLT produced accurate results. In future work, replace HMLT hybrid method with other high optimal methods, such as natural inspiring methodologies.

ACKNOWLEDGMENT

This work was supported by the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R192), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors would like to thank VIGNAN'S INSTITUTE OF MANAGEMENT AND TECHNOLOGY FOR WOMEN for their useful advices.

REFERENCES

[1] T. D. Rathnaweera, W. Wu, Y. Ji, and R. P. Gamage, ''Understanding injection-induced seismicity in enhanced geothermal systems: From the coupled thermo-hydro-mechanical-chemical process to anthropogenic earthquake prediction,'' *Earth-Sci. Rev.*, vol. 205, Jun. 2020, Art. no. 103182.

[2] G. Cremen and C. Galasso, ''Earthquake early warning: Recent advances and perspectives,'' *Earth-Sci. Rev.*, vol. 205, Jun. 2020, Art. no. 103184.

[3] K. Cao and Q. Huang, ''Geo-sensor(s) for potential prediction of earthquakes: Can earthquake be predicted by abnormal animal phenomena?'' *Ann. GIS*, vol. 24, no. 2, pp. 125–138, Apr. 2018.

[4] N. G. Granin, N. A. Radziminovich, M. De Batist, M. M. Makarov, V. V. Chechelnitcky, V. V. Blinov, I. A. Aslamov, R. Y. Gnatovsky, J. Poort, and S. G. Psakhie, ''Lake Baikal's response to remote earthquakes: Lake-level fluctuations and near-bottom water layer temperature change,'' *Mar. Petroleum Geol.*, vol. 89, pp. 604–614, Jan. 2018.

[5] X. Yan, Z. Shi, P. Zhou, H. Zhang, and G. Wang, ''Modeling earthquake-induced spring discharge and temperature changes in a fault zone hydrothermal system,'' *J. Geophys. Res., Solid Earth*, vol. 125, no. 7, Jul. 2020, Art. no. e2020JB019344.

[6] V. Shukla, V. Chauhan, N. Kumar, and D. Hazarika, ''Assessment of Rn-222 continuous time series for the identification of anomalous changes during moderate earthquakes of the Garhwal Himalaya,'' *Appl. Radiat. Isot.*, vol. 166, Dec. 2020, Art. no. 109327.

[7] K. Kawabata, T. Sato, H. A. Takahashi, F. Tsunomori, T. Hosono, M. Takahashi, and Y. Kitamura, ''Changes in groundwater radon concentrations caused by the 2016 Kumamoto earthquake,'' *J. Hydrol.*, vol. 584, May 2020, Art. no. 124712.

[8] X. Wang, N. Hu, H. Fukunaga, and Z. H. Yao, ''Structural damage identification using static test data and changes in frequencies,'' *Eng. Struct.*, vol. 23, no. 6, pp. 610–621, Jun. 2001.

[9] P. H. Kirkegaard and A. Rytter, ''Use of a neural network for damage detection and location in a steel member,'' Dept. Building Technol. Struct. Eng., Aalborg Univ., Aalborg, Denmark, Tech. Rep. 41, 1992, vol. R9245.

[10] N. Leelawat, A. Suppasri, I. Charvet, and F. Imamura, ''Building damage from the 2011 Great East Japan tsunami: Quantitative assessment of influential factors,'' *Natural Hazards*, vol. 73, no. 2, pp. 449–471, Sep. 2014.

[11] A. Suppasri, K. Pakoksung, I. Charvet, C. T. Chua, N. Takahashi, T. Ornthammarath, P. Latcharote, N. Leelawat, and F. Imamura, ''Load-resistance analysis: An alternative approach to tsunami damage assessment applied to the 2011 Great East Japan tsunami,'' *Natural Hazards Earth Syst. Sci.*, vol. 19, no. 8, pp. 1807–1822, Aug. 2019.

[12] N. Leelawat, A. Suppasri, O. Murao, and F. Imamura, ''A study on the influential factors on building damage in Sri Lanka during the 2004 Indian Ocean tsunami,'' *J. Earthq. Tsunami*, vol. 10, no. 2, Jun. 2016, Art. no. 1640001.

[13] A. Suppasri, E. Mas, I. Charvet, R. Gunasekera, K. Imai, Y. Fukutani, Y. Abe, and F. Imamura, ''Building damage characteristics based on surveyed data and fragility curves of the 2011 Great East Japan tsunami,'' *Natural Hazards*, vol. 66, no. 2, pp. 319–341, Mar. 2013.

[14] I. Charvet, I. Ioannou, T. Rossetto, A. Suppasri, and F. Imamura, ''Empirical fragility assessment of buildings affected by the 2011 Great East Japan tsunami using improved statistical models,'' *Natural Hazards*, vol. 73, no. 2, pp. 951–973, Sep. 2014.

[15] A. Suppasri, K. Pakoksung, I. Charvet, N. Takahashi, T. Ornthammarath, and F. Imamura, ''New building damage assessment method based on numerical tsunami simulation and analytical force estimation: Case study of the 2011 Tohoku tsunami,'' *J. Jpn. Soc. Civil Eng., Ser.*, vol. 74, no. 2, pp. 451–456, 2018.

[16] P. Debnath, P. Chittora, T. Chakrabarti, P. Chakrabarti, Z. Leonowicz, M. Jasinski, R. Gono, and E. Jasinska, ''Analysis of earthquake forecasting in India using supervised machine learning classifiers,'' *Sustainability*, vol. 13, no. 2, p. 971, Jan. 2021.

[17] P. Lu, S. Chen, and Y. Zheng, ''Artificial intelligence in civil engineering,'' *Math. Problems Eng.*, vol. 2012, pp. 1–22, Dec. 2012.

[18] K. Saengtabtim, N. Leelawat, J. Tang, W. Treeranurat, N. Wisittiwong, A. Suppasri, K. Pakoksung, F. Imamura, N. Takahashi, and I. Charvet, ''Predictive analysis of the building damage from the 2011 Great East Japan tsunami using decision tree classification related algorithms,'' *IEEE Access*, vol. 9, pp. 31065–31077, 2021.

[19] Y. Bao, Y. Guo, and H. Li, ''A machine learning–based approach for adaptive sparse time–frequency analysis used in structural health monitoring,'' *Struct. Health Monitor.*, vol. 19, no. 6, pp. 1963–1975, Nov. 2020.

[20] Y. Bao and H. Li, ''Machine learning paradigm for structural health monitoring,'' *Struct. Health Monitor.*, vol. 20, no. 4, pp. 1353–1372, 2021.

[21] H. Adeli, ''Neural networks in civil engineering: 1989–2000,'' *Comput.-Aided Civil Infrastruct. Eng.*, vol. 16, no. 2, pp. 126–142, Mar. 2001.

[22] Z. Meng, G. Li, X. Wang, S. M. Sait, and A. R. Yildiz, ''A comparative study of metaheuristic algorithms for reliability-based design optimization problems,'' *Arch. Comput. Methods Eng.*, vol. 28, no. 3, pp. 1853–1869, May 2021.

[23] V. Dave, S. Singh, and V. Vakharia, ''Diagnosis of bearing faults using multi fusion signal processing techniques and mutual information,'' *Indian J. Eng. Mater. Sci.*, vol. 27, no. 4, pp. 878–888, 2021.

[24] S. Ahadzadeh and M. R. Malek, ''Earthquake damage assessment in three spatial scale using Naive Bayes, SVM, and deep learning algorithms,'' *Appl. Sci.*, vol. 11, no. 20, p. 9737, Oct. 2021.

[25] H. Zhou, X. Wang, and R. Zhu, ''Feature selection based on mutual information with correlation coefficient,'' *Appl. Intell.*, vol. 52, no. 5, pp. 5457–5474, 2021.

[26] J. M. Steppe, K. W. Bauer, and S. K. Rogers, ''Integrated feature architecture selection,'' *IEEE Trans. Neural Netw.*, vol. 7, no. 4, pp. 1007–1014, Jul. 1996.

[27] N. Kwak and C.-H. Choi, ''Input feature selection for classification problems,'' *IEEE Trans. Neural Netw.*, vol. 13, no. 1, pp. 143–159, Jan. 2002.

[28] A. Srivastava, A. Asati, and M. Bhattacharya, ''Fast hybrid rough-set theoretic fuzzy clustering technique with application to multispectral image segmentation,'' in *Proc. 1st Int. Conf. Intell. Interact. Technol. Multimedia*, Dec. 2010, pp. 126–129.

[29] D. K. Srivastava, K. S. Patnaik, and L. Bhambhu, ''Data classification: A rough-SVM approach,'' *Contemp. Eng. Sci.*, vol. 3, no. 2, pp. 77–86, 2010.

[30] U. Dackermann, J. Li, and B. Samali, ''Damage index method for damage identification utilising artificial neural networks,'' in *Proc. 9th Int. Conf. Motion Vib. Control*, 2009, pp. 15–18.

[31] Z. Zuowei and M. Lili, ''The integrated methodology of rough set theory and artificial neural network for safety assessment on construction sites,'' in *Proc. 3rd Int. Conf. Intell. Human-Mach. Syst. Cybern.*, Aug. 2011, pp. 105–109.

[32] M. V. V. Rao and A. Chaparala, ''Analysis of classification technique for prediction of damages levels in building-structures,'' *Int. J. Adv. Sci. Technol.*, vol. 29, no. 5, pp. 822–842, 2020.

[33] A. Chaparala, ''A building damage classification framework for feature subset selection using rough set with mutual information,'' *Solid State Technol.*, vol. 63, pp. 498–509, pp. 498–509, Feb. 2020.

[34] M. V. V. Rao and A. Chaparala, ''An efficient data mining technique for structural strength monitoring system,'' *Ingénierie Des Syst. d Inf.*, vol. 26, no. 2, pp. 237–243, Apr. 2021.

[35] M. V. V. Rao, ''A machine learning-based damage prediction techniques for structural health monitoring,'' *Turkish J. Comput. Math. Educ. (TURCOMAT)*, vol. 12, no. 2, pp. 3392–3405, Apr. 2021.

[36] M. Rao and A. Chaparala, ''A novel feature-based SHM assessment and predication approach for robust evaluation of damage data diagnosis systems,'' *Wireless Personal Commun.*, vol. 124, no. 4, pp. 3387–3411, 2022.

[37] C. Cortes and V. Vapnik, ''Support vector machine,'' *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

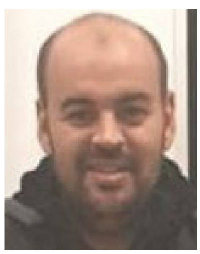
[38] H. Hothu and A. Mita, "Damage detection method using support vector machine and first three natural frequencies for shear structures," vol. 2013, pp. 104–112, Jun. 2013.
[39] R. Ghazali, A. J. Hussain, and P. Liatsis, "Dynamic ridge polynomial neural network: Forecasting the univariate non-stationary and stationary trading signals," *Exp. Syst. Appl.*, vol. 38, no. 4, pp. 3765–3776, Apr. 2011.
[40] Y. Chu, H. T. C. Pedro, and C. F. M. Coimbra, "Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning," *Sol. Energy*, vol. 98, pp. 592–603, Dec. 2013.
[41] Y. Chu, H. T. C. Pedro, L. Nonnenmacher, R. H. Inman, Z. Liao, and C. F. M. Coimbra, "A smart image-based cloud detection system for intrahour solar irradiance forecasts," *J. Atmos. Ocean. Technol.*, vol. 31, no. 9, pp. 1995–2007, Sep. 2014.
[42] Y. Chu, B. Urquhart, S. M. I. Gohari, H. T. C. Pedro, J. Kleissl, and C. F. M. Coimbra, "Short-term reforecasting of power output from a 48 MWe solar PV plant," *Sol. Energy*, vol. 112, pp. 68–77, Feb. 2015.
[43] A. Ramdas, B. Poczos, A. Singh, and L. Wasserman, "An analysis of active learning with uniform feature noise," 2015, *arXiv:1505.04215*.
[44] C. Y. Kao and S.-L. Hung, "Detection of structural damage via free vibration responses generated by approximating artificial neural networks," *Comput. Struct.*, vol. 81, nos. 28–29, pp. 2631–2644, Nov. 2003.
[45] M. Sahin and R. A. Shenoi, "Quantification and localisation of damage in beam-like structures by using artificial neural networks with experimental validation," *Eng. Struct.*, vol. 25, no. 14, pp. 1785–1802, Dec. 2003.
[46] B. Xu, Z. Wu, G. Chen, and K. Yokoyama, "Direct identification of structural parameters from dynamic responses with neural networks," *Eng. Appl. Artif. Intell.*, vol. 17, no. 8, pp. 931–943, Dec. 2004.
[47] J. J. Lee, J. W. Lee, J. H. Yi, C. B. Yun, and H. Y. Jung, "Neural networks-based damage detection for bridges considering errors in baseline finite element models," *J. Sound Vibrat.*, vol. 280, nos. 3–5, pp. 555–578, Feb. 2005.
[48] S. Ahadzadeh, M. Valadanzouj, S. Sadeghian, and S. Ahmadi, "Detection of damaged buildings after an earthquake using artificial neural network algorithm," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 37, pp. 369–372, Jan. 2008.
[49] B. Efron, "Estimating the error rate of a prediction rule: Improvement on cross-validation," *J. Amer. Stat. Assoc.*, vol. 78, no. 382, pp. 316–331, 1983.
[50] P. Krizek, "Feature selection: Stability, algorithms, and evaluation," Ph.D. thesis, Czech Tech. Univ. Prague, Prague, Czechia, 2008.
[51] T. L. Bailey and C. Elkan, "Estimating the accuracy of learned concepts," in *Proc. Int. Joint Conf. Artif. Intell.*, 1993, pp. 1–6.
[52] D. K. Sharma, P. Shrivastava, and S. Garg, "Utilizing word embedding and linguistic features for fake news detection," in *Proc. 9th Int. Conf. Comput. for Sustain. Global Develop. (INDIACom)*, Mar. 2022.
[53] D. P. Yadav, A. S. Jalal, and V. Prakash, "Human burn depth and grafting prognosis using ResNeXt topology based deep learning network," *Multimedia Tools Appl.*, vol. 81, no. 13, pp. 18897–18914, May 2022.
[54] S. C. Agrawal and A. S. Jalal, "Distortion-free image dehazing by superpixels and ensemble neural network," *Vis. Comput.*, vol. 38, no. 3, pp. 781–796, Mar. 2022.
[55] D. P. Yadav, A. Sharma, S. Athithan, A. Bhola, B. Sharma, and I. B. Dhaou, "Hybrid SFNet model for bone fracture detection and classification using ML/DL," *Sensors*, vol. 22, no. 15, p. 5823, Aug. 2022.

VISHNU VARDHANA RAO M. received the M.Tech. degree in computer science and engineering from Acharya Nagarjuna University, Guntur, Andhra Pradesh, India, and the Ph.D. degree from the Department of Computer Science and Engineering, Acharya Nagarjuna University. His Ph.D. research focuses on data mining methods for applications in building damage prediction. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Vignan's Institute of Management and Technology for Women, Kondapur, Ghatkesar, Medchal, India. He has ten years of teaching experience and has published seven technical articles in international scholarly journals. He has also published papers in many conferences, such as Springer and IEEE. He was also associated with reputed journals, such as *Transactions on Emerging Telecommunications Technologies* (Wiley), *Artificial Intelligence Review* (Springer), *The Journal of Supercomputing* (Springer), and IEEE ACCESS, in 2020 and 2021, as a Reviewer. He is highly interested in social activity and providing free-of-cost education to underprivileged students.

APARNA CHAPARALA received the master's degree from the Department of Computer Science and Systems Engineering, Andhra University, in 2004, and the Ph.D. degree from Jawaharlal Nehru Technological University, Hyderabad, in 2014. Her Ph.D. dissertation was titled, "Enhanced Classification Techniques for Brain–Computer Interface Data." She is currently a Professor with the Department of Computer Science and Engineering, R. V. R. & J. C. College of Engineering, and guiding research scholars in the areas of machine learning, cyber forensics, and databases. She has around 20 research publications in indexed journals and has presented papers at conferences. Her research interests include data mining, databases, information security, and computer networks.

PRAPHULA KUMAR JAIN received the B.E. degree in computer science and engineering from the Faculty of Computer Science and Engineering, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, Madhya Pradesh, India, and the M.Tech. degree in computer science and engineering from the Indian Institute of Technology (ISM) Dhanbad, Jharkhand, India. His Ph.D. research focuses on machine learning methods for various applications. He is currently a Ph.D. Researcher with the Department of Computer Science and Engineering, Indian Institute of Technology (ISM) Dhanbad. He has three years of teaching experience and has published many technical articles in international scholarly journals, such as *ACM Transactions on Asian and Low-Resource Language Information Processing*, *Computer Science Review* (Elsevier), *Computers & Electrical Engineering* (Elsevier), *Journal of Ambient Intelligence and Humanized Computing* (Springer), *The Journal of Supercomputing* (Springer), *Multimedia Tools and Applications*, *Wireless Personal Communications* (Springer), *International Arab Journal of Information Technology* (CCIS), *SN Applied Sciences* (Springer), and *OPSEARCH* (Springer). He has also presented papers in many conferences, such as RAIT and GUCON. He is also associated with reputed journals, such as *Transactions on Emerging Telecommunications Technologies* (Wiley), *Artificial Intelligence Review* (Springer), *The Journal of Supercomputing* (Springer), IEEE ACCESS, *Future Foods* (Elsevier), *AI and Ethics*, and *AI & Society*, and conferences, such as FSDM, GLOBECOM 2020, and GLOBECOM 2021, as a Reviewer. He is highly interested in social activity and providing free-of-cost education to underprivileged students.

HANEN KARAMTI received the bachelor's degree in computer science and multimedia from the Higher Institute of Computer Science and Multimedia of Sfax, University of Sfax, Tunisia, the master's degree in computer science and multimedia from University of Sfax, in 2011, and the joint Ph.D. degree in computer science from the National Engineering School of Sfax, Tunisia, in 2017, in cooperation with the University of La Rochelle, France, and Hanoi University, Vietnam. She is currently an Assistant Professor with Princess Nourah Bint Abdul Rahman University, Riyadh, Saudi Arabia. Her research interests include information retrieval, multimedia systems, image retrieval, health informatics, big data, and data analytics.



WALID KARAMTI was born in Sfax, Tunisia, in 1985. He received the bachelor's, master's, and Ph.D. degrees in computer science from the University of Sfax, Tunisia, in 2008, 2010, and 2015, respectively.

From 2015 to 2019, he was an Assistant Professor with the University of Monastir, Tunisia. He is currently an Assistant Professor with Qassim University, Saudi Arabia. He is the author/coauthor of several papers in international conferences and journals. His current research interests include formal methods for the analysis of behaviors in real-time systems and social networks.