

ACADEMIA OAMENILOR
DE ȘTIINȚĂ DIN ROMÂNIA



Raport științific nr. 2 privind proiectul

Platformă DIGItală pentru esTimarea nivelului de dEgradaRe

seismică a clădiRilor utilizând tehnici de mAchine learning

DIGITERRA

Domeniu științific : Științe Geonomice

Estimarea degradărilor structurale utilizând algoritmi de machine learning

Cuvinte cheie : algoritmi de machine learning ; analiza dinamică
neliniară ; degradare structurală ; digitalizare ; reziliență

Conf.dr.ing CALOFIR Vasile
Cpt.dr.ing. MUNTEANU Ruben-Iacob
As.dr.ing. SIMOIU Mircea-Stefan

- Director de proiect
- Membru în echipa de cercetare
- Membru în echipa de cercetare

Universitatea Națională de Știință și Tehnologie
POLITEHNICA BUCUREȘTI

Facultatea de Automatică și Calculatoare



Estimarea degradărilor structurale utilizând algoritmi de machine learning

CUPRINS

I	Introducere și scop	1
II	Starea cercetării	1
III	Metoda propusă	2
III-A	Generarea modelelor structurale și centralizarea datelor relevante	2
III-B	Procedura de simulare dinamică neliniară	3
III-C	Propunere framework estimare	5
III-D	Algoritmii de estimare	5
III-E	Procedura de evaluare a performanțelor	6
III-F	Procesul de selecție a caracteristicilor	6
IV	Studiu de caz	6
IV-A	Generarea bazei de date cu modelele structurale și răspunsul seismic	6
IV-B	Analiza algoritmilor de învățare automată - rezultate și discuții	9
V	Concluzii	12
	Referințe	13

LISTĂ DE FIGURI

1	Structuri de clădiri regulate și neregulate	3
2	Procedura de generare a modelelor structurale și centralizarea datelor relevante	3
3	Procedura de calcul a răspunsului dinamic neliniar	4
4	Framework-ul de simulare propus, subliniind maniera secvențială în care s-au evidențiat principalele caracteristici utilizat în procesul de învățare automată, precum și principalele performate utilizate în procesul de evaluare pentru estimarea eficientă a indicelui de degradare	5
5	Procesul de selecție a caracteristicilor bazat pe utilizarea algoritmilor de învățare automată	7
6	Spectrul de răspuns al celor 20 accelerații utilizate ca intrări seismice	7
7	Geometria clădirilor regulate și neregulate	7
8	Distribuția primei perioade naturale pentru clădirile regulate	8
9	Distribuția indicelui de degradare pentru clădirile regulate	8
10	Distribuția accelerației seismice maxime (scalată) pentru încărcarea pe direcția X	8
11	Distribuția accelerației seismice maxime (scalată) pentru încărcarea pe direcția Y	8
12	Distribuția primei perioade naturale pentru clădirile neregulate	8
13	Distribuția indicelui de degradare pentru clădirile neregulate	8
14	Ierarhia importanței caracteristicilor, având în vedere metoda propusă în Figura 5. Rezultatele pot fi reproduse folosind Jupyter Notebook disponibil la [1].	10
15	Comparație între indicii de degradare estimați și indicii de degradare reali pentru fiecare punct de date asociat activității seismice a clădirii din setul de date de testare, pentru fiecare pereche formată din modelul de învățare automată (Regresie liniară, KNN și KNN optimizat) și setul de caracteristici. Rezultatele pot fi reproduse folosind Jupyter Notebook disponibil la [2]	11
16	Reprezentarea rezultatelor în spațiu obiectiv 2x2, cu punctul ideal în originea sistemului de coordonate. În ultimele două diagrame, linia albastră leagă soluțiile dominante, raportate la indicatorii utilizați pentru evaluare. Ploturile pot fi reproduse folosind Jupyter Notebook disponibil la [2]	12

LISTĂ DE TABELE

I	Setul de date ce descriu caracteristicile clădirilor regulate	14
II	Setul de date ce descriu caracteristicile clădirilor neregulate	15
III	Setul de date cu principalele caracteristici ale inputurilor seismice	16

Estimarea degradărilor structurale utilizând algoritmi de machine learning

Rezumat—Metodologiile convenționale de proiectare structurală folosesc adesea tehnici de analiză elastică, cum ar fi metoda forței statice echivalente și metoda răspunsului spectral. În timp ce aceste metode sunt cunoscute pentru simplitatea și eficiența lor numerică, ele se dovedesc a fi ineficiente în estimarea degradărilor structurale cauzate de forțele seismice. În plus, utilizarea analizei dinamice neliniare pentru estimarea degradărilor structurale reprezintă o procedură dificilă și complicată, generând dificultăți în proiectarea structurală. Prin urmare, obiectivul acestui raport de cercetare este de a prezenta o metodologie inovatoare pentru evaluarea degradărilor structurale seismice ale structurilor în cadre. Acest lucru implică utilizarea algoritmilor de machine learning, care au fost antrenați și testați pe un set mare de date generate folosind o procedură de simulare eficientă numeric de actualitate. Pentru a estima cu precizie indicele de degradare structurală Park-Ang se folosesc algoritmi de învățare automată de tip regresie liniară și de tip KNN (K-Nearest Neighbors).

I. INTRODUCERE ȘI SCOP

În zilele noastre este foarte important ca în regiunile seismice clădirilor să fie realizate cu un grad ridicat de rezilientă. Astfel, structurile trebuie proiectate nu doar pentru a rezista la impactul imediat al cutremurelor ci având ca obiectiv principal și reducerea pierderilor umane și economice, astfel încât, să se garanteze funcționalitatea continuă sau realuarea funcționării în cel mai scurt timp.

Un element cheie în atingerea rezilienței seismice a structurilor este utilizarea analizei dinamice neliniare. Acest tip de simulare oferă o estimare precisă a modului în care clădirile răspund la forțele seismice, ținând cont totodată și de complexitatea condițiilor reale. Spre deosebire de analiza liniară, simulările dinamice neliniare pot surprinde comportarea inelastică a elementelor structurale, oferind o perspectivă cuprinzătoare asupra degradărilor structurale așteptate în urma acțiunilor seismice.

Strategiile tradiționale de proiectare structurală folosesc adesea metode de analiză elastică precum abordarea forței statice echivalente și metoda răspunsului spectral. În timp ce aceste metode sunt mai simple și mai puțin complexe din punct de vedere numeric, pot prezenta riscul de a nu surprinde comportamentul complex al structurilor supuse acțiunii forțelor seismice. În același timp, deși analiza dinamică neliniară poate fi mai precisă, se dovedește totuși a fi o sarcină destul de complexă care poate crea dificultăți pentru majoritatea proiectanților de structuri.

Aici intervine utilitatea algoritmilor de învățare automată. Odată ce acești algoritmi sunt antrenați pe seturi extinse de date care să includă evaluări ale degradărilor pentru o gamă diversă de modele structurale, aceștia pot fi folosiți pentru a estima cu

precizie degradările structurale pentru alte clădiri cunoscând doar caracteristicile generale clădirilor și ale cutremurelor la care acestea au fost supuse. Acest lucru nu numai că simplifică activitatea de realizare a analizelor structurale, ci oferă și o perspectivă valoroasă care poate ajuta în mod semnificativ inginerii structuristi și companiile de profil în proiectarea unor clădiri reziliente.

Obiectivul principal al acestui raport este de a prezenta o abordare inovatoare pentru estimarea degradărilor structurale, folosind algoritmi de învățare automată antrenați pe date generate prin intermediul unei proceduri numerice de simulare, dezvoltată anterior de către echipa de cercetare și diseminată în articole precum [3] și [4]. Algoritmul de învățare automată este remarcabil prin capacitatea sa de a estima cu precizie nivelul degradărilor structurale, utilizând pentru aceasta un set limitat de parametri de intrare.

II. STAREA CERCETĂRII

În ultimii ani, domeniul dinamicii structurilor a suferit schimbări revoluționare datorită apariției tehnologiilor de învățare automată. În trecut, acesta era dominat de modele numerice tradiționale și metode analitice. În schimb, în momentul de față, în acest domeniu și-au făcut loc cu succes algoritmi de învățare automată ce sunt capabili să gestioneze cu precizie și eficiență ridicată, sisteme complexe și neliniare. Prin urmare, acest capitol prezintă succint aplicațiile recente în care sunt prezentați algoritmi de machine learning folosiți pentru estimarea degradărilor structurale cauzate de incidența cutremurelor.

Unul dintre primele studii care pot fi găsite în literatura de specialitate [5] prezintă o metodă generală de predicție a degradărilor induse de acțiunile seismice folosind rețele neuronale artificiale (ANN). Pentru antrenarea ANN-urilor au fost folosite un set de 2200 modele de cadre din beton armat (RC) 2D ce au variat în topologie, rigiditate, rezistență și amortizare. Dintre acestea, au fost folosite în faza de testare 619 clădiri. Dinamica răspunsului structural a fost simulată utilizând analiza neliniară FEM, în timp ce pentru estimarea amplitudinii degradărilor structurale au fost utilizați indicii Park și Ang. În consecință, folosind un model de tip ANN, a fost stabilită o corespondență între proprietățile structurii și a solului și indicii de degradare. Rezultatele au arătat că modelul ANN a fost capabil să prezică cu exactitate nivelul degradărilor structurale. O altă lucrare importantă [6], a vizat dezvoltarea unei metode, bazate pe inteligența artificială, care să fie utilizată în analiza performanțelor seismice ale clădirilor 3D din beton armat (RC) din Turcia. Astfel, au fost selectate pentru studiu un total de 66 RC de structuri cuprinse între 4 și

10 etaje, reprezentând clădirile tipice RC din Turcia. Nivelul de performanță seismică a acestor clădiri a fost determinat folosind analiza elastică și nivelurile de performanță cu 4 grade, specificate în Codul de proiectare seismică din Turcia-2007 (*TEC – 2007*). Prin urmare, a fost antrenat un model ANN cu scopul de a reproduce nivelul de performanță al fiecărei clădiri folosind ca mărimi de intrare 19 parametri care descriu mișcarea seismică și respectiv caracteristicile clădirii. Rezultatele au arătat că algoritmul poate ajunge la o precizie de până la 80%. O altă abordare pentru a investiga dacă ANN-urile pot fi utilizate eficient pentru a estima starea de degradare seismică a clădirilor RC a fost propusă de [7]. Acest studiu tratează problema atât ca o problemă de aproximare a funcției, cât și ca o problemă de recunoaștere a modelelor, folosind rețele Feedforward Perceptron multistrat. Setul de date pentru antrenarea modelului a constat în 30 de clădiri RC supuse la 65 de mișcări seismice înregistrate, în timp ce degradările structurale au fost evaluate efectuând analize neliniare ale clădirilor și pe baza valorii maxime a deplasării relative dintre etaje. Studiul a explorat, de asemenea, impactul diferiților parametri de configurare a modelului ANN asupra robusteții predicției și a testat capacitățile de generalizare ale modelului. Cea mai importantă concluzie a lucrării este că ANN-urile pot estima într-un mod robust starea de degradare seismică a clădirilor RC după cutremur. Mai mult, acești autori au investigat numărul și combinația de parametri seismici prin care se poate realiza o predicție optimă pentru starea de degradare a clădirilor RC [8]. Într-o lucrare mai recentă, [9] s-a investigat potențialul utilizării unei metode alternative pentru estimarea degradărilor seismice în clădirile din zidărie. Astfel, s-a efectuat o analiză comparativă a două abordări diferite: metoda indicelui de vulnerabilitate și o abordare inovatoare bazată pe utilizarea ANN-urilor. Rezultatele analitice obținute prin strategiile menționate au fost comparate cu seturi de date reale privind degradările structurale colectate după cutremurul de 1998 din Azore. Rezultatele au arătat că ANN-ul a fost capabil să ofere estimări mult mai bune decât cele realizate cu abordarea cu indicii de vulnerabilitate. În același timp, din rezultatele oferite de modelul ANN a fost obținută o nouă expresie analitică pentru estimarea degradărilor structurale.

Deși aceste studii sunt extrem de complexe, există unele probleme care nu au fost încă abordate. Acestea includ utilizarea unor seturi de date mai extinse care captează un spectru mai larg de modele structurale cu neregularități atât de plan, cât și de elevație, luând în considerare, de asemenea și accelerogramele dintr-o anumită zonă geografică, cu scopul de a include în timpul analizei neliniare și caracteristicile unice ale solului respectiv. În plus, având în vedere seturile de caracteristici menționate mai sus, un aspect interesant de investigat poate fi și determinarea unui set de date adecvate cu scopul de fi utilizate în estimarea indicelui de degradare folosind algoritmi de machine learning, cum ar fi regresia liniară sau K-nearest neighbors (KNN).

Acest raport urmărește continuarea eforturilor de cercetare depuse până acum prin introducerea unei noi perspective asupra modului în care tehnicile de învățare automată pot fi utilizate pentru a estima indicii de degradare a structurilor în cadre din beton armat ce sunt supuse la acțiuni seismice.

Principala contribuție a autorilor este crearea unui instrument numeric complex, care să permită cercetătorilor și inginerilor structuristi să dezvolte rutine foarte precise bazate pe învățarea automată pentru estimarea degradărilor structurale ale clădirilor supuse mișcărilor seismice din zone geografice specifice. Elementele cheie ale acestei platforme numerice sunt utilizarea unei proceduri rapide de simulare a răspunsului dinamic neliniar și implementarea unei tehnici eficiente de machine learning care să ia în considerare atât procedura de selecție a caracteristicilor considerate, cât și o optimizare hiperparametrică a algoritmului de învățare automată.

III. METODA PROPUȘĂ

După cum sa menționat anterior, obiectivul principal al raportului este de a oferi un instrument numeric care facilitează estimarea degradărilor structurale folosind algoritmi de tip machine learning. Acest instrument este compus din două părți principale. În primul rând, a fost dezvoltată o rutină numerică pentru generarea unui set mare de modele structurale diverse și, ulterior, realizarea de simulări dinamice neliniare cu acestea pentru estimarea degradărilor seismice. Scopul acestor simulări este realizarea unui set extins de date care să conțină informații reprezentative despre modelele structurale, mișcarea seismică și răspunsul structural. În etapa a doua a cercetării, setul de date generat anterior, a fost utilizat pentru antrenarea a doi algoritmi avansați de învățare automată cu scopul de a estima cu exactitate degradările structurale.

A. Generarea modelelor structurale și centralizarea datelor relevante

Pentru generarea modelelor structurale a fost utilizată Metoda Elementelor Finite. Astfel, pentru reprezentarea cu acuratețe a comportamentului tridimensional al grinzilor și al stâlpilor, a fost folosit un element finit specific de tip bară cu 2 noduri și 6 grade de libertate pentru fiecare nod. În consecință, fiecărui element structural în parte i-au fost asociate matrice locale de masă și rigiditate care sunt asamblate cu scopul de a descrie masa și rigiditatea globală întregului model. Matricea de amortizare \mathbf{C} a fost construită folosind abordarea Reylegh, fiind astfel defintă ca proporțională cu matricea maselor \mathbf{M} și cea de rigiditate \mathbf{K} .

Mai mult, pentru a acoperi un număr mare de tipologii de clădiri pentru setul de date, a fost dezvoltată o subrutină de generare automată a modelelor structurale. În consecință, pentru fiecare analiză dinamică neliniară se construiește câte un model unic prin alegerea aleatorie a: dimensiunii totale în plan, numărului de travee, numărului de deschidei, numărul de etaje, înălțimea etajului, capacitățile la înconvoiere ale grinzilor și stâlpilor, încărcarea uniform distribuită, modulul lui Young al sistemului structural și coeficientul de amortizare critică al betonului armat. Pentru a construi modele structurale cu caracteristici plauzibile, aceste valori au fost selectate în mod arbitrar dintr-un set de intervale atent alese. Mai mult, este important de menționat că modelele generate se încadrează în două categorii distincte: cele cu geometrie regulată și cele cu geometrie neregulată. Modelele regulate prezintă o formă dreptunghiulară pe fiecare etaj (Figura 1 a), în timp ce modelele

neregulate prezintă un design în formă de L , cu o configurație dreptunghiulară parțială pe etajele superioare Figura 1 b).

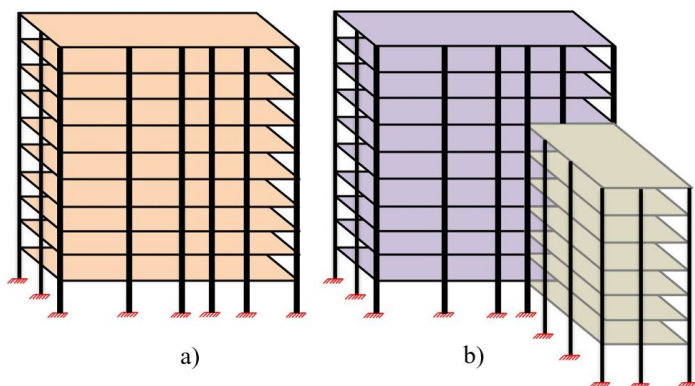


Figura 1: Structuri de clădiri regulate și neregulate

În același timp, pentru a ne se asigura că modelele structurale generate aleatoriu sunt construite rațional, a fost implementat un proces suplimentar de filtrare pe două niveluri. Primul nivel, este folosit înainte de începerea analizei dinamice neliniare și este legat de comportarea liniară al modelului. Acesta presupune acceptarea drept plauzibile numai acele modele pentru care prima perioadă naturală de vibrație T_1 se încadrează în intervalul $0,07 \cdot no_story < T_1 < 0,2 \cdot no_story$ unde no_story este numărul de etaje al clădirii. Al doilea nivel de filtrare este aplicat după terminarea simulării dinamice și este legat de comportamentul nelinier al modelului. Acesta implică acceptarea răspunsului structural ca fiind plauzibil numai dacă indicii de degradare Park și Ang este sub valoarea de 2,5. Secvența procedurală este ilustrată în Figura 2.

După fiecare simulare, este generat răspunsul structural al clădirii prin agregarea datelor obținute de la fiecare pas de timp discret din timpul analizei dinamice neliniare. În acest sens, trebuie menționat că în măsura în care se dorește păstrarea întregii evoluții a variabilelor structurale de răspuns obținute în timpul simulării, cum ar fi deplasările, vitezele, momentele încovoietoare și parametrii aferenți, este necesar aproximativ 1,5 gigabytes de memorie pe hard disk. Având în vedere numărul mare de analize dinamice efectuate, devine foarte important să se reducă din ocuparea memoriei. În consecință, după simulare, se rețin selectiv doar parametrii esențiali care descriu modelul structural și răspunsul nelinier. În tabelele I, II și III este prezentată distribuția setului de date ce a fost utilizat în antrenarea și testarea algoritmilor de învățare automată.

B. Procedura de simulare dinamică neliniară

O abordare comună în ingineria civilă pentru modelarea comportamentului nelinier implică utilizarea unor metode simplificate. De exemplu, în cazul modelelor în cadre cu mai multe etaje rezistente la moment, plasticitatea este concentrată la nivelul fiecărei articulații plastice modelând astfel comportamentul nelinier prin relații forță-deplasare. Deși au fost propuse în literatura de specialitate mai multe modele complexe care utilizează plasticitatea distribuită pentru a surprinde

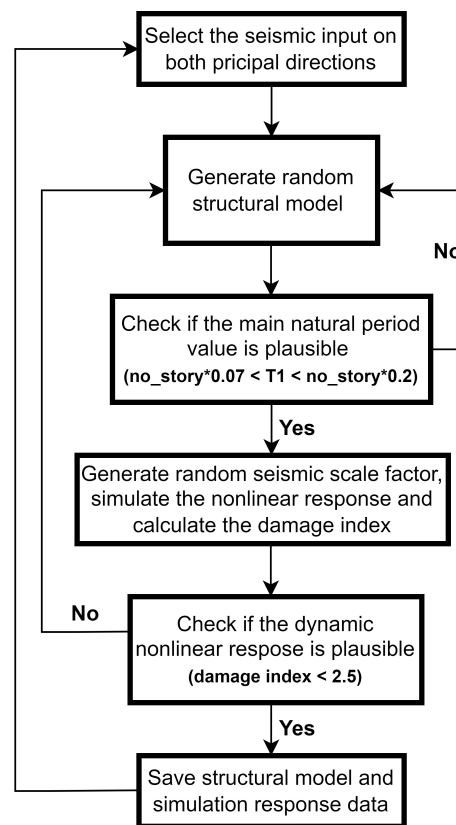


Figura 2: Procedura de generare a modelelor structurale și centralizarea datelor relevante

comportamentul inelastic, modelele simplificate s-au dovedit a fi mai practice datorită eficienței numerice și formulărilor acestora relativ concise. Aceste modele simplificate au arătat că pot surprinde caracteristicile relevante ale răspunsului structural cu un nivel de precizie identic sau doar marginal redus.

Deși există diferite abordări numerice posibile pentru modelarea comportamentului nelinier folosind articulații plastice, raportul de față folosește o metodă recent apărută, și anume, Metoda Analogiei Forțelor (FAM).

Prima lucrare publicată referitoare la FAM este [10] și prezintă conceptele fundamentale, inclusiv principiile de aplicare în domeniul simulării comportamentului dinamic nelinier al structurilor. De-a lungul anilor, metoda a evoluat și a fost dezbătută pe larg, inclusiv în cărți precum [11] și [12], oferind astfel detalierea principiilor care au stat la baza FAM și exemple ilustrative în acest sens. Mai mult, autorii prezentului raport au adus contribuții semnificative prin extinderea aplicării FAM de la modele bidimensionale la modele tridimensionale cu comportament histeretic complex, așa cum este detaliat în referințele [4] și [13].

În cercetările recente, [14] a folosit FAM pentru efectuarea a 9600 de analize dinamice neliniare unidimensionale pentru a dezvolta o procedură automată de evaluare a fragilității seismice a structurilor 3D din cadre din beton armat. Durata totală pentru simularea celor 9600 de analize s-a ridicat la

aproape 128 ore, rezultând un timp mediu de procesare de aproximativ 40 secunde per analiză. Prin urmare, autorii acestei lucrări au demonstrat că FAM poate fi utilizat cu succes pentru a efectua seturi mari de analize dinamice neliniare într-un interval de timp foarte rezonabil. În consecință, pornind de la cadrul numeric dezvoltat de acești autori, raportul de față introduce îmbunătățiri care implică luarea în considerare a efectelor seismice pe ambele direcții principale și utilizarea modelelor structurale cu o gamă mai largă de caracteristici geometrice. În plus, în timp ce studiul anterior a efectuat 9600 de analize folosind 24 de modele structurale, cercetarea de față a implicat analize dinamice neliniare pentru 59569 de modele structurale distincte, care au servit drept date de antrenare și testare pentru algoritmi de învățare automată.

Ideea principală a FAM este de a adăuga în ecuația de mișcare un nou vector de valori necunoscute care să țină cont de deplasările inelastice $\mathbf{u}''(t)$ ale modelului structural Eq. 1. Acest lucru asigură că matricea de rigiditate a structurii, determinată la începutul analizei, rămâne constantă pe tot parcursul calculului inelastice. Atunci când este combinată cu reprezentarea în spațiului stărilor, FAM devine o abordare robustă, precisă și eficientă pentru simularea răspunsului dinamic neliniar.

$$\mathbf{M} \cdot \ddot{\mathbf{u}}(t) + \mathbf{C} \cdot \dot{\mathbf{u}}(t) + \mathbf{K} \cdot \mathbf{u}(t) = -\mathbf{M} \cdot \mathbf{h} \cdot \ddot{\mathbf{a}}_g(t) + \mathbf{K} \cdot \mathbf{u}''(t) \quad (1)$$

Procedura de simulare seismică, așa cum este prezentată în Figura 3, constă din două operații principale: În primul rând, folosind datele pasului anterior, se calculează vectorul deplasărilor \mathbf{u}_{k+1} prin utilizarea procedurii de integrare explicite din Ec. 2, dezvoltat de [12]. În al doilea rând, pentru a calcula vectorul deplasării plastice $\mathbf{u}''(t)$, se calculează valorile de încercare (trial) pentru momentul încovoietor \mathbf{m}_t și vectorul ce conține incrementul de rotire inelastică $\Delta\theta''$. Aceste cantități descriu nivelurile de inconvoiere și deformare inelastica de la nivelul fiecărei articulații plastice. În același timp, este calculat vectorul forțelor axiale \mathbf{n}_{k+1} ca o funcție a vectorului de deplasare de la pasul curent \mathbf{u}_{k+1} . Apoi, valorile din vectorul \mathbf{m}_t sunt comparate individual cu capacitățile de încovoiere ale articulațiilor plastice. Trebuie menționat că în cazul grinzilor s-a folosit o articulație plastică de tip moment încovoietor, în timp ce pentru stâlpi s-a considerat interacțiunea dintre forța axială și încovoierea pe ambele direcții principale folosind o suprafață de capacitate $3D$ ($n - m_y - m_z$). Dacă încovoierea depășește limita elastică, valorile sunt limitate la capacitate, efortul de depășire fiind considerat un increment de deformație plastică necunoscut ce trebuie de determinat. Ulterior este utilizat un sistem cu nph (numărul total de **plastic hinge**) ecuații și nph necunoscute pentru a determina valorile actualizate \mathbf{m}_t și $\Delta\theta''$. În continuare, valorile din \mathbf{m}_t sunt din nou comparate cu capacitatea pentru a evalua dacă a avut loc vreo redistribuire de eforturi de încovoiere inelastica. Dacă există valori noi care depășesc limita elastică, \mathbf{m}_t este trimis înapoi la rutina de capacitate, în caz contrar, sunt calculate momentele finale de încovoiere \mathbf{m}_{k+1} și rotirile plastice θ'' . În cele din urmă, vectorul deplasărilor plastice este calculat și trimis în formula de integrare explicită pentru a determina

vectorul de deplasare de la următorul pas de timp. În esență, calculul vectorului deplasărilor plastice se bazează pe alegerea unei serii de valori de încercare ce sunt ajustate iterativ până la atingerea echilibrului.

$$\mathbf{z}_{k+1} = e^{\mathbf{A} \cdot \Delta t} \cdot \mathbf{z}_k + e^{\mathbf{A} \cdot \Delta t} \cdot \mathbf{H} \cdot \Delta t \cdot \ddot{\mathbf{a}}_{gk} + e^{\mathbf{A} \cdot \Delta t} \cdot \mathbf{F}_p^c \cdot \Delta t \cdot \mathbf{u}_k'' \quad (2)$$

unde

$$\mathbf{z}_{k+1} = \begin{bmatrix} \mathbf{u}_{k+1} \\ \dot{\mathbf{u}}_{k+1} \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} \mathbf{O}_n & \mathbf{I}_n \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix} \quad (3)$$

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{u}_k \\ \dot{\mathbf{u}}_k \end{bmatrix} \quad \mathbf{F}_p^c = \begin{bmatrix} \mathbf{O}_n \\ -\mathbf{M}^{-1}\mathbf{K} \end{bmatrix}$$

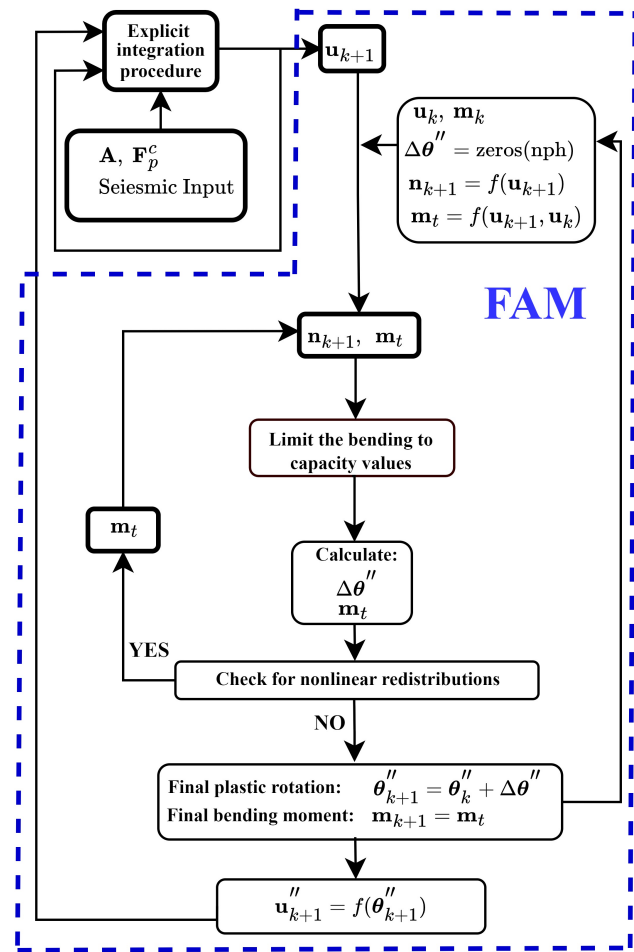


Figura 3: Procedura de calcul a răspunsului dinamic neliniar

Pentru realizarea analizei dinamice neliniare este utilizată o bază de date de accelerații seismice. În acest proces, platforma numerică dezvoltată selectează inputurile seismice care sunt utilizate pentru fiecare analiză în parte.

C. Propunere framework estimare

Având în vedere că obiectivul principal al raportului este estimarea indicelui de degradare pe baza unui set extins de caracteristici, în Figura 4 este reprezentată structura framework-ului de estimare.

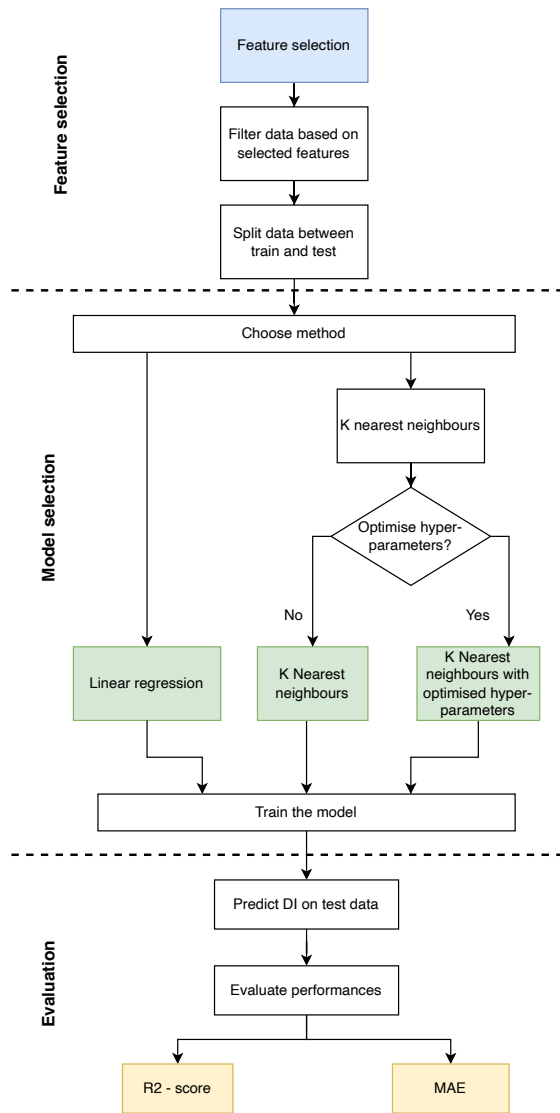


Figura 4: Framework-ul de simulare propus, subliniind maniera secvențială în care s-au evidențiat principalele caracteristici utilizat în procesul de învățare automată, precum și principalele performate utilizate în procesul de evaluare pentru estimarea eficientă a indicelui de degradare

În faza inițială se consideră că avem un set complet de date, caracterizat de multiple caracteristici. Concret, fiecare punct al setului de date este reprezentat de o clădire, cu caracteristicile sale și principalii parametri ai mișcării seismice la care modelul structural a fost supus.

Având în vedere acest set de date, un prim obiectiv este

alegerea caracteristicilor cu cel mai mare grad de corelație cu indicele de degradare. Pentru a realiza acest lucru a fost implementat un proces de selecție a caracteristicilor pornind de la un set cuprinzător de caracteristici ce au fost selectate corespunzător pe baza cunoștințelor specializate de inginerie structurală. Acest pas este conceput pentru a identifica mai multe seturi de caracteristici cheie pe baza criteriilor alese corespunzător, oferind astfel perspective multiple în direcția identificării unei metodologii optime.

Următorul pas este împărțirea datelor într-un set de date de antrenare și unul de testare. Odată ce modelul de învățare automată este calibrat pe setul de date de antrenare, acesta poate fi utilizat eficient pentru a estima indicele de degradare pe baza caracteristicilor de intrare ale clădirilor și ale accelerației seismice prezente în setul de date de testare. Având în vedere setul amplu de date generat descris în secțiunea anterioară, nu apare nicio restricție specială cu privire la alegerea procentelor în care sunt selectate seturile de date pentru antrenare și testare.

În acest raport considerăm ca și metode de estimare algoritmi precum: o regresie liniară tipică din spectrul modelelor liniare și K-Nearest-Neighbours (KNN), care este un model mai degrabă potrivit pentru estimarea relațiilor mai complexe sau neliniare. Raționamentul din spatele acestei alegeri este derivat dintrun obiectiv secundar al acestui raport, și anume de a prezenta o analiză comparativă între performanțele modelelor liniare tipice (de exemplu regresie liniară) în raport cu cele ale modelelor mai complexe.

În cele din urmă, luând în considerare indicatorii tipici de evaluare, cum ar fi scorul R^2 sau eroarea medie absolută (MAE), sunt comparate predicțiile modelelor alese realizate cu datele de testare. Obiectivul acestei evaluări finale este de a oferi o explicație cuprinzătoare a modului în care diferite seturi de caracteristici, atunci când sunt analizate împreună cu diferite modele, afectează metricile menționate mai sus. Aceasta, la rândul său, determină acuratețea modelelor în estimarea indicelui de degradare, precum și determinarea unui set de caracteristici de dimensiune minimă care trebuie luat în considerare în realizarea acestui obiectiv de predicție.

D. Algoritmii de estimare

Primul algoritm pe care îl folosim este regresia liniară. Un model de regresie liniară reprezintă un instrument utilizat pe scară largă pentru estimarea unei variabile dependente y pe baza unei variabile (sau caracteristici) independente x , ținând cont totodată că există o relație liniară între variabilele dependente și independente.

Forma generală a modelului este:

$$y = xa + b \quad (4)$$

unde a este termenul de interceptare și b este panta.

Acest model simplist este potrivit pentru aplicații simple de estimare, unde este suficientă o singură caracteristică pentru a furniza informații pentru estimarea variabilei dependente. Totuși, având în vedere contextul acestui raport în care se urmărește estimarea indicelui de degradare pe baza multiplelor

caracteristici ale clădirii și ale accelerației seismice, o opțiune mai potrivită este reprezentată următorul model:

$$y = XA + b \quad (5)$$

unde X este un set de variabile (sau caracteristici) independente și A reprezintă un vector de coeficienți corespunzători interceptului. Acest tip de model reprezintă un prim pas în problema de estimare a indicelui de degradare seismică și poate reprezenta o soluție convenabilă, datorită simplității sale.

Cu toate acestea, având în vedere relația neliniară evidentă dintre caracteristici și indicele de degradare, am utilizat ulterior un algoritm alternativ și anume algoritmul K-nearest neighbors (KNN) [15]. Deși algoritmul KNN este utilizat în mod tradițional pentru probleme de clasificare, unde este folosit pentru gruparea entităților analoge (vecini) în funcție de distanțele euclidiene dintre caracteristicile acestora, algoritmul poate fi folosit și în probleme de estimare.

În contextul aplicației prezentate în acest raport, problema nu poate fi strict caracterizată ca o problemă de clasificare, ci mai degrabă ca una de regresie. Adaptarea algoritmului KNN în scopuri de regresie este concepută pentru a funcționa parcurgând următorii pași secvențiali:

- 1) Se alege numărul de vecini 'K' de utilizat în procesul de estimare (o decizie fundamentală prezentă și în varianta de utilizarea a algoritmului pentru operația de clasificare).
- 2) Pentru un nou set de caracteristici ale modelului structural și ale accelerației seismice, pentru care scopul este estimarea indicelui de degradare, se calculează distanța euclidiană dintre noul vector caracteristici și toate celelalte din setul de date selectat.
- 3) Sunt identificate cele mai apropiate puncte de date 'K' din apropierea noului vector de caracteristici.
- 4) Se calculează valoarea medie a indicelui de degradare a acestor vecini 'K' și se atribuie valoarea medie ca indice de degradare estimat noului vector de caracteristici.

Alegerea acestui algoritm pentru estimarea indicelui de degradare rezidă în mod specific din simplitatea sa în rezolvarea problemelor de estimare care includ mai multe componente neliniare.

E. Procedura de evaluare a performanțelor

Pentru evaluarea algoritmilor de învățare automată au fost utilizați 2 indicatori principali: eroarea medie absolută și scorul R^2 .

Eroarea medie absolută (MAE) reprezintă eroarea medie dintre valorile estimate ale indicelui de degradare seismică și valorile reale.

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

Având în vedere scara utilizată pentru cuantificarea indicelui, eroarea MAE scăzută ne arată că indicele de degradare estimat este relativ apropiat de cel adevărat.

Scorul R^2 este un indicator de performanță foarte utilizat care ne arată cât de mult din varianța variabilei dependente

vine ca o consecință a varianței unei variabile independente. De obicei, un scor R^2 de 1 arată că estimarea datelor reale este foarte bună, în timp ce un scor R^2 de 0,4 sau mai mic poate arăta că alegerea modelului sau a caracteristicilor implicate în procesul de predicție ar putea fi îmbunătățită.

Întrucât în setul de date utilizat indicele de degradare oscilează între 0 și 2.5, considerăm că acești doi indicatori pot reprezenta o metodă adecvată pentru evaluarea algoritmilor de învățare automată. Deși fiecare indicator posedă caracteristici distincte (de exemplu, MAE poate fi susceptibil la valori aberante), este imperativ să se ia în considerare și numărul de caracteristici ca potențial indicator. Această considerație subliniază importanța setului de date necesar pentru fiecare scenariu de estimare. Printr-o astfel de abordare, poate fi oportun de a identifica ce seturi de caracteristici produc performanțe satisfăcătoare, utilizând doar o parte dintre variabilele independente. Această constatare este deosebit de avantajoasă în contextul acestui raport, deoarece nu toți parametrii structurii pot fi ușor măsurabili.

F. Procesul de selecție a caracteristicilor

Procesul de selecție a caracteristicilor relevante pentru procedura de estimare se bazează pe analiza singulară a fiecărui atribut, în raport cu cele două modele de predicție menționate anterior. Mai exact, așa cum este descris în figura 5, alegem fiecare caracteristică din setul de caracteristici, aplicăm un algoritm pentru a estima indicele de degradare bazat exclusiv pe acea caracteristică izolată și, în final, comparăm valoarea estimată de algoritmul de învățare automată cu valoarea reală a indicelui de degradare. Această analiză comparativă folosește două metrici ca indicatori ai performanței: coeficientul de determinare (scor R^2) și eroarea medie absolută (MAE).

IV. STUDIU DE CAZ

Această secțiune este dedicată prezentării eficienței framework-ului de simulare propus. Acesta constă din doi pași cheie. În primul rând, sunt efectuate simulări dinamice neliniare pentru a genera o bază de date cuprinzătoare care conține date despre modelele structurale și răspunsul seismic al acestora. Ulterior, folosind această bază de date este realizat procesul de antrenare și testare a algoritmilor de învățare automată, regresie liniară și K-nearest neighbors, cu scopul de a oferi estimări precise ale indicelui de degradare pentru modelele structurale analizate.

A. Generarea bazei de date cu modelele structurale și răspunsul seismic

Pentru a genera baza de date care descrie modelele structurale și răspunsul seismic al acestora, au fost utilizate trei calculatoare din laboratoarele Universității Naționale de Știință și Tehnologie POLITEHNICA BUCUREȘTI. Simulările s-au întins pe o perioadă de trei luni, implicând un total de 59569 de analize. Simulările dinamice neliniare au fost efectuate utilizând un set de 20 de mișcări seismice obținute de la trei cutremure de adâncime medie din Vrancea (România) care au avut loc în 1977, 1986 și 1990. Acest set de date a fost ales în

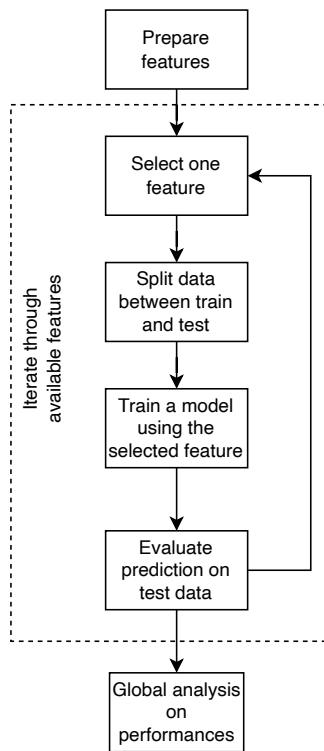


Figura 5: Procesul de selecție a caracteristicilor bazat pe utilizarea algoritmilor de învățare automată

mod special pentru a reprezenta condițiile locale ale solului din București [16]. În figura 6 este reprezentat spectrul de răspuns al accelerației pentru fiecare dintre aceste înregistrări. Astfel, se poate observa că cele mai pronunțate amplificări spectrale sunt concentrate în intervalul perioadei medii. Prin urmare, pentru fiecare simulare în parte este selectată ca și intrare seismică de-a lungul direcțiilor principale X și Y, câte o pereche de accelerograme. Mai mult, înainte de începerea analizei, aceste accelerograme trec printr-o etapă de preprocesare, în care le este atribuit aleatoriu un factor de scalare din intervalul [1-4]. În figurile 10 și 11 sunt prezentate pentru fiecare dintre direcțiile principale, accelerațiile de vârf ale solului asociate mișcărilor seismice selectate pentru efectuarea simulărilor.

De asemenea, în figurile 8- 13 este ilustrată în corelație cu figura 7 o scurtă prezentare a modelelor structurale și a rezultatelor simulării. Pentru claritate, cele 59569 de simulări sunt împărțite în două părți: unul cuprinzând rezultate pentru modele de geometrie regulată (29997 de simulări) și celălaltă parte pentru modele de geometrie neregulată (29572 simulări).

În ceea ce privește modelele cu geometrie regulată, figura 8 ilustrează faptul că majoritatea modelelor (92, 5%) au prima perioadă naturală cuprinsă între 0,4s și 1,3s, ceea ce este în concordanță cu numărul de etaje, considerate în faza de generare aleatorie. Mai mult, în Figura 9 este prezentată distribuția valorii indicelui de degradare Park Ang. Se poate observa că 55,57% dintre modele sunt sub valoarea 0,4, care

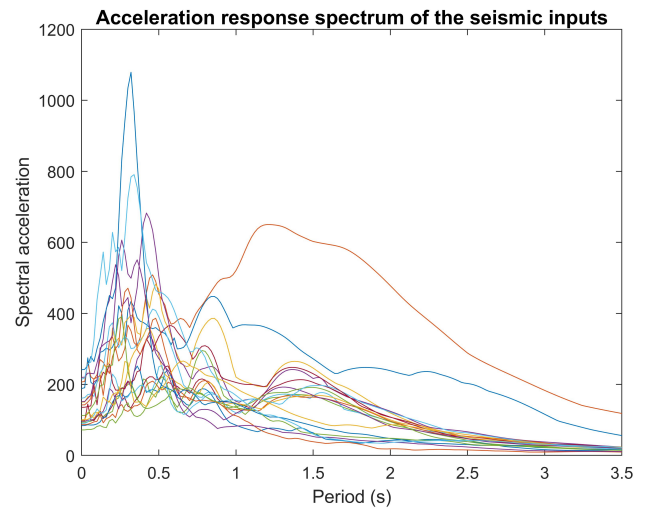


Figura 6: Spectrul de răspuns al celor 20 accelerații utilizate ca intrări seismice

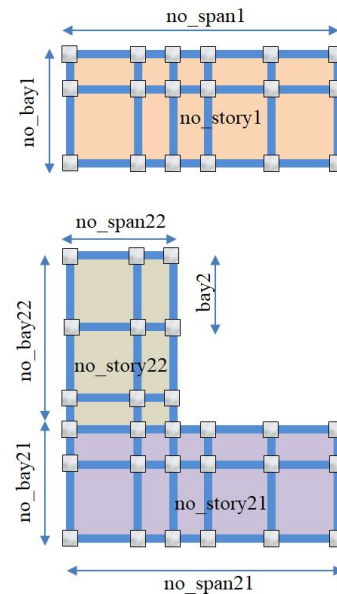


Figura 7: Geometria clădirilor regulate și neregulate

este considerată limita degradărilor reparabile, în timp ce 14,37% depășesc valoarea 1 corespunzătoare colapsului. Mai mult, în Figura 12 este prezentată distribuția primei perioade naturale a modelelor neregulate. Se poate observa că aceste structuri sunt mai puțin flexibile decât cele regulate, aproape toate valorile fiind situate sub 2s. În ceea ce privește distribuția indicelui de degradare, se poate observa că rezultatele sunt similare cu cele obținute pentru modelele regulate, 61,18% dintre valori fiind sub 0,4 în timp ce 10,72% peste 1.

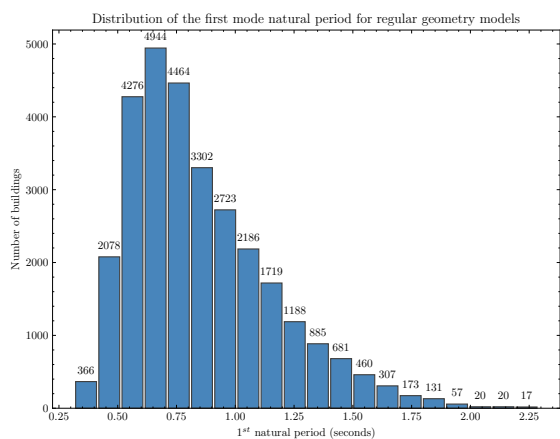


Figura 8: Distribuția primei perioade naturale pentru clădirile regulate

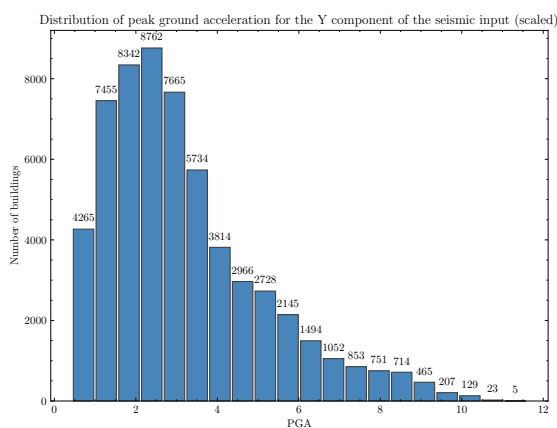


Figura 11: Distribuția accelerației seismice maxime (scalată) pentru încărcarea pe direcția Y

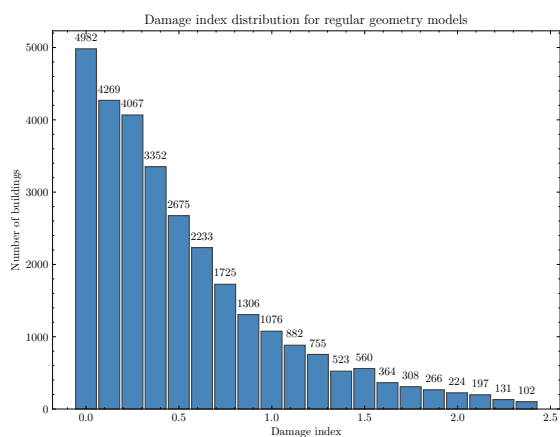


Figura 9: Distribuția indicelui de degradare pentru clădirile regulate

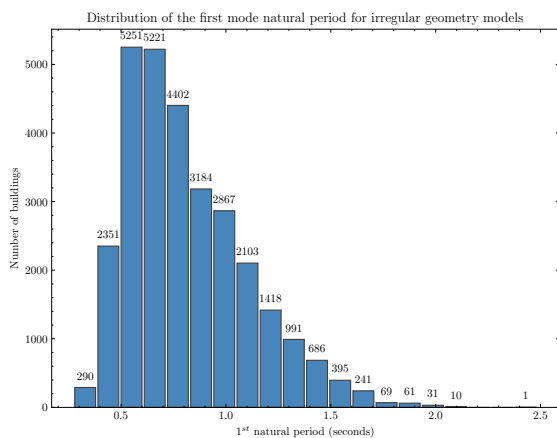


Figura 12: Distribuția primei perioade naturale pentru clădirile neregulate

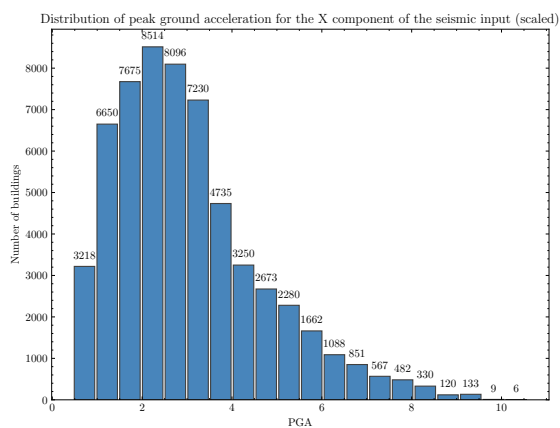


Figura 10: Distribuția accelerației seismice maxime (scalată) pentru încărcarea pe direcția X

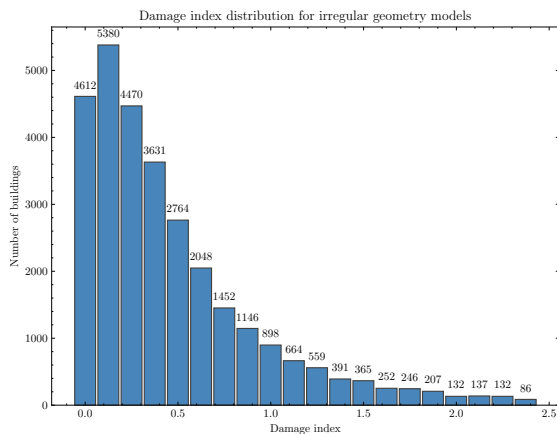


Figura 13: Distribuția indicelui de degradare pentru clădirile neregulate

B. Analiza algoritmilor de învățare automată - rezultate și discuții

Pornind de la metodologia descrisă în figura 4, prima fază în obținerea unui model de machine learning asociat unui set adecvat de caracteristici începe cu procesul de selecție a caracteristicilor. Setul inițial de caracteristici propus în raportul de față, constă din următorii parametri: Fund_Freq1, Fund_Freq2, PGA_of_the_recording_scale_1, PGA_of_the_recording_scale_2, b_Hieght, dim_x, dim_y, b_st, h_st, b_gr, h_gr, E, MstY, MstX, Mgr, Lshape, bay2, nr_span22, nr_bay22, nr_story22, T1, T2 și T3. Distribuția valorilor acestor parametri este detaliată în tabelele I, II și III. Trebuie subliniat că aceste caracteristici au fost alese pe baza cunoștințelor de specialitate de inginerie structurală ale autorilor. În plus, indicele de degradare este desemnat ca variabilă dependentă. Setul de date este constituit ulterior din X puncte de date, ce cuprinde caracteristicile menționate mai sus și o valoare corespunzătoare a indicelui de degradare.

În acest scop, a fost aplicat setului de date metoda de selecție a caracteristicilor prezentată în figura 5. Astfel, a fost evaluată fiecare combinație a unei caracteristici cu un algoritm de învățare automată, obținând astfel un scor $R2$ și o valoare MAE . Algoritmii utilizați la acest pas sunt un model general de regresie liniară și algoritmul KNN cu 5 vecini. Scopul este de a determina importanța fiecărei caracteristici în cadrul procesului de estimare, prin utilizarea indicatorilor de performanță – $R2$ și MAE – în evaluarea setului de date de testare. În plus, pentru acest pas inițial, setul de date este împărțit în două componente: un set de date de antrenare care reprezintă 80% din totalul datelor și un set de date pentru testare care conține restul punctelor de date. Rezultatele sunt prezentate în figura 5.

Evaluând performanțele modelului de regresie liniară, se pare că nicio caracteristică particulară nu poate forma o relație liniară puternică cu indicele de degradare. În afară de caracteristicilor accelerației seismice, toate celelalte caracteristici sunt asociate cu valori scăzute ale scorului $R2$ și cu valori ridicate ale scorului MAE , rezultând astfel o eroare semnificativă în estimare.

În schimb, la examinarea metricilor de performanță asociate algoritmului KNN, se observă cum o colecție de nouă caracteristici exercită o influență substanțială asupra procesului de estimare. Această influență este subliniată atât de scorul $R2$, cât și de valorile MAE . În lumina acestor constatări, sunt propuse pentru fazele ulterioare ale analizei învățării automate următoarele șase seturi de caracteristici :

- setul "toate caracteristicile" - care cuprinde fiecare caracteristică propusă. Acest set cuprinzător este un punct de referință inițial critic de luat în considerare.
- setul "cele mai bune 9 caracteristici" - care este compus din primele nouă caracteristici, determinate în procesul de selecție a caracteristicilor și evaluate pe baza scorului $R2$.
- Setul "cele mai bune 8 caracteristici", care reflectă comportamentul setului anterior, dar este limitat la primele opt caracteristici.
- Setul "cele mai bune 7 caracteristici", care, la fel ca pre-

decesorii săi, este limitat la primele șapte caracteristici.

- Setul "cele mai bune 6 caracteristici", care se limitează la cele mai bune șase caracteristici.
- Setul "cele mai bune 2 caracteristici", care include numai cele mai bune două caracteristici.

Pe lângă găsirea celui mai bun model, seturile de mai sus sunt alese având în vedere obiectivul de a obține un număr minim de caracteristici care oferă o platformă bună pentru estimarea indicelui de degradare.

Faza ulterioară a analizei implică pregătirea modelelor utilizând seturile de caracteristici obținute anterior. Respectând metodologia descrisă în figura 4, sunt folosite trei tipuri de modele pentru a estima indicele de degradare: un model de regresie liniară, un model KNN cu cinci vecini și o versiune optimizată a modelului KNN. Acest model KNN optimizat este determinat printr-o căutare sistematică care explorează o combinație de număr de vecini (3, 5, 11, 15) și o metodă uniformă sau ponderată de cuantificare a distanței dintre vecini. Metoda uniformă implică faptul că toți vecinii contribuie în mod egal la predicție, în timp ce în cazul metodei ponderate vecinii mai apropiați au o influență mai mare asupra predicției decât vecinii mai îndepărtați.

În consecință, pentru fiecare pereche formată dintr-un set de caracteristici și un algoritm de învățare automată, este ilustrată în Figura 15 o analiză comparativă a valorilor reale versus cele estimate. În cadrul fiecărui grafic, punctele individuale simbolizează perechile corespunzătoare punctelor de date din setul de date de testare; aceste perechi constau din indicele de degradare real și omologul său estimat. Pentru a crea o referință vizuală a unei estimări perfecte, se trasează o linie punctată roșie din origine la un unghi de 45 de grade, reprezentând linia de-a lungul căreia indicele de degradare real este egal cu indicele de degradare estimat.

Astfel, se pot observa reprezentările din figura 15 din două perspective: perspectiva tipului de model și perspectiva mărimii setului de caracteristici implicat în antrenarea modelului.

Din perspectiva tipurilor de model, se poate observa că cele mai bune performanțe sunt date de algoritmi KNN, versiunea optimizată generând un număr semnificativ de puncte de date estimate care sunt egale cu cele reale. Un alt rezultat care poate fi observat este legat de performanța regresiei liniare, ceea ce arată că modelele de regresie liniară nu pot fi utilizate în acest context pentru estimarea indicelui de degradare. Grafic, acest aspect este subliniat de răspândirea punctelor de date în raport cu axa de referință de 45 grade. S-ar putea considera acest rezultat relativ așteptat, având în vedere importanța fiecărei caracteristici individuale în procesul de selecție a caracteristicilor folosind modelul de regresie liniară (Figura 14).

În cele din urmă, este prezentată în figura 16 o analiză mai aprofundată, în care aceleași perechi constând dintr-un set de caracteristici și un algoritm de estimare sunt descrise într-o reprezentare de 2 câte 2 a trei indicatori: scorul $R2$, MAE și raportul de utilizare a caracteristicilor (FUR). Raportul de utilizare a caracteristicilor cuantifică proporția de caracteristici utilizate în procesul de estimare în raport cu numărul total de caracteristici disponibile.

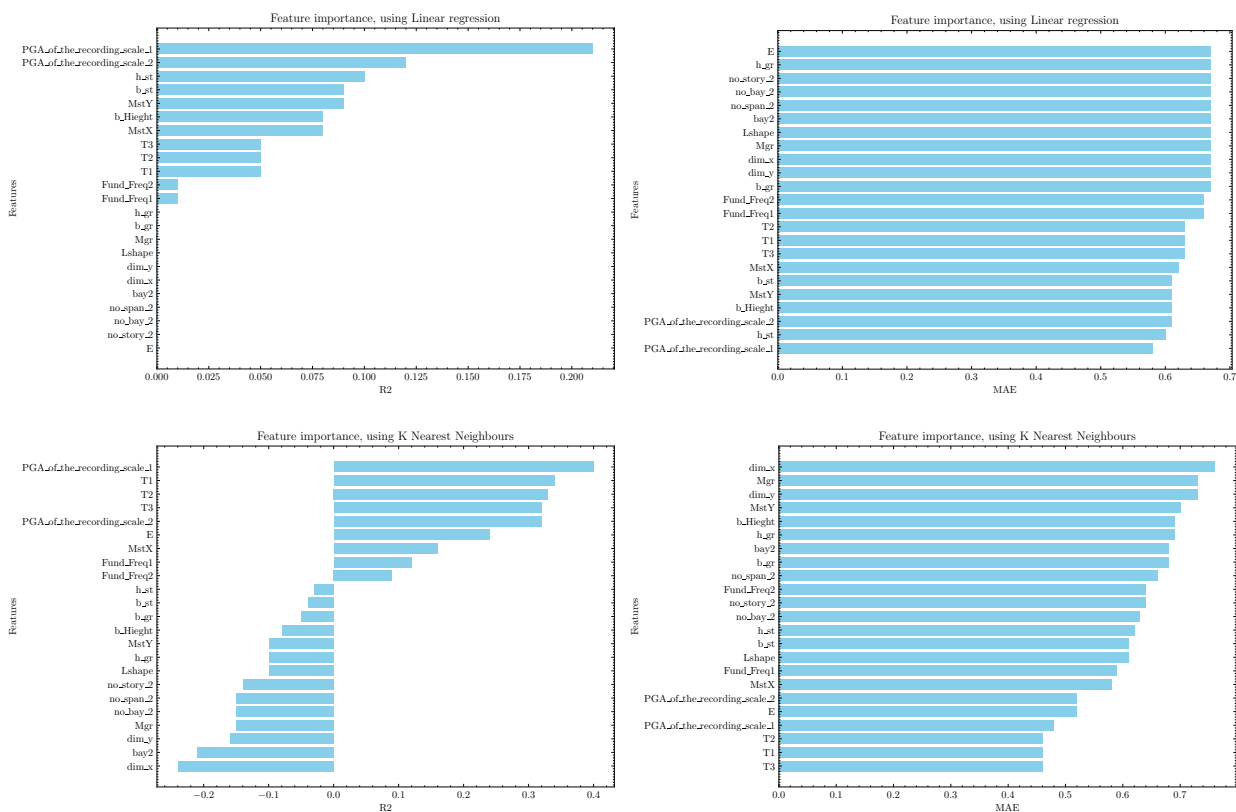


Figura 14: Ierarhia importanței caracteristicilor, având în vedere metoda propusă în Figura 5. Rezultatele pot fi reproduse folosind Jupyter Notebook disponibil la [1].

Această metodă grafică își propune să contextualizeze performanța fiecărui scenariu în raport cu un punct de referință optim, desemnat de originea fiecărui sistem de coordonate. De exemplu, într-o diagramă cu eroarea absolută medie și raportul de utilizare a caracteristicilor, originea reprezintă scenariul ipotetic de realizare a predicției perfecte fără a utiliza nicio caracteristică. Acest tip de reprezentare este similară cu reprezentarea grafică Pareto utilizată în optimizare ([17]), unde este de asemenea trasată o diagramă pentru a evidenția punctele dominante în ceea ce privește ambele performanțe.

După cum putem vedea, în prima diagramă care arată scenariile din perspectiva scorului MAE și $R2$, cele mai bune modele și seturi de caracteristici sunt afișate cel mai aproape de originea graficului. Acolo, vedem că modelul KNN optimizat aplicat tuturor celor 9 caracteristicilor oferă cel mai bun scor $R2$ și cel mai mic MAE . Ulterior, versiunea optimizată KNN oferă în general cele mai bune performanțe, urmată îndeaproape de modelul KNN cu 5 vecini. Această observație întărește alegerea unui model de tip KNN pentru estimarea indicelui de degradare și validează eficacitatea procesului de selecție a caracteristicilor cu cel mai mare impact în estimarea indicelui de degradare din întreg setul de date.

Graficele ulterioare oferă informații suplimentare despre performanță, încorporând, de asemenea, raportul de utilizare

a caracteristicilor. În acest caz, cele mai bune soluții sunt interconectate printr-un grafic care subliniază dominanța față de celelalte soluții. Mai mult, aceste diagrame ajută utilizatorul final să determine cel mai potrivit set de caracteristici și cel mai potrivit model conform nevoilor sale. De exemplu, dacă nu sunt constrângerți în obținerea datelor și sunt obținute ușor toate cele nouă caracteristici, atunci ar fi recomandat pentru estimare modelul KNN optimizat, deoarece asigură o performanță foarte bună în ceea ce privește scorul MAE și $R2$. Această pereche de caracteristici-model este marcat în legendă printr-un pătrat albastru poziționat în partea stânga a plotului. Alternativ, dacă nu sunt disponibile multe caracteristici, cea mai bună alegere în ceea ce privește setul de caracteristici ar putea fi plasată în partea dreaptă a plotului, spre soluția care cuprinde doar două caracteristici și performanță de estimare relativ scăzută.

În general, prin folosirea procesului de selecție a caracteristicilor propus și folosind algoritmul KNN optimizat, se poate estima cu precizie indicele de degradare, pe baza caracteristicilor clădirii și ale intrării seismice asociate. Chiar dacă modelele liniare precum regresia liniară nu reprezintă o soluție de estimare adecvată, modelul KNN este capabil să obțină performanțe foarte bune atât în ceea ce privește scorul $R2$ (egal cu 1) cât și MAE (aproximativ 0.18).

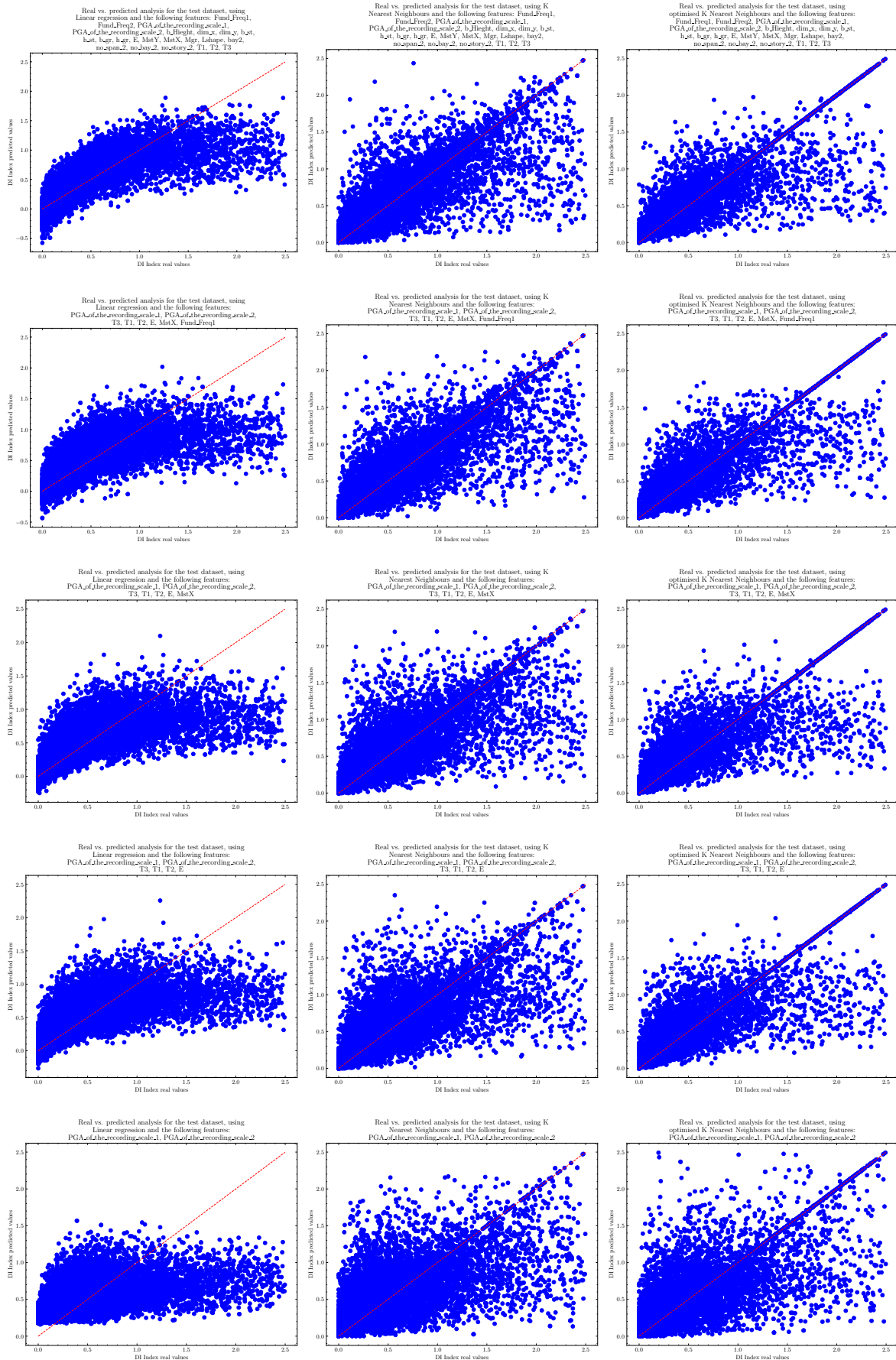


Figura 15: Comparație între indicii de degradare estimați și indicii de degradare reali pentru fiecare punct de date asociat activității seismice a clădirii din setul de date de testare, pentru fiecare pereche formată din modelul de învățare automată (Regresie liniară, KNN și KNN optimizat) și setul de caracteristici. Rezultatele pot fi reproduse folosind Jupyter Notebook disponibil la [2]

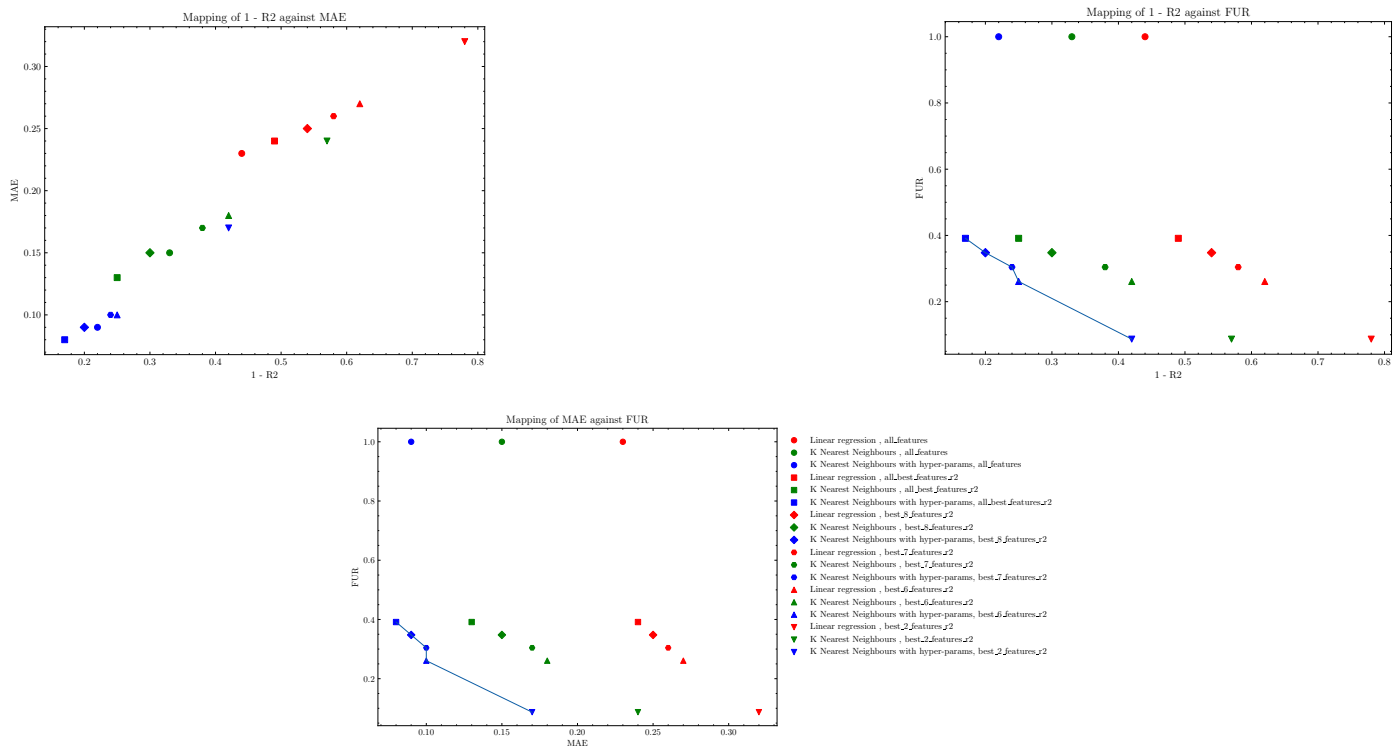


Figura 16: Reprezentarea rezultatelor în spațiu obiectiv 2x2, cu punctul ideal în originea sistemului de coordonate. În ultimele două diagrame, linia albastră leagă soluțiile dominante, raportate la indicatorii utilizați pentru evaluare. Ploturile pot fi reproduse folosind Jupyter Notebook disponibil la [2]

V. CONCLUZII

Scopul acestui raport este de a introduce o metodologie inovatoare pentru evaluarea degradărilor seismice ale structurilor în cadre rezistente la momente. Aceasta implică aplicarea algoritmilor de învățare automată care sunt antrenați și testați pe un set mare de date generate printr-o procedură de simulare eficientă. Algoritmii de învățare automată folosesc diferite abordări de regresie pentru a estima indicele de degradare structurală Park-Ang.

Setul de date utilizat în această raport este compus din 59569 de modele structurale unice și răspunsul seismic neliniar asociat acestora, exprimat sintetic prin indicele de degradare Park - Ang. În comparație cu studiile prezentate în [5] și [6], cercetarea noastră le depășește atât în dimensiunea setului de date, cât și în complexitate. Modelele structurale luate în considerare sunt generate aleatoriu în funcție de criteriile de plauzibilitate specifice și au structuri geometrice atât regulate, cât și neregulate, oferind astfel o reprezentare cuprinzătoare a diverselor tipuri de clădiri întâlnite în practica curentă.

Mai mult, este propus un framework de învățare automată pentru a determina atât caracteristicile cele mai potrivite, cât și un algoritm de estimare adecvat, capabil să estimeze cu precizie indicele de degradare pe baza accelerației seismice și a caracteristicilor clădirii. Procedura presupune selectarea unui set minim de caracteristici importante și alegerea celui

mai bun algoritm pentru estimarea indicelui de degradare. Rezultatele arată că un algoritm K-Nearest Neighbors cu hiper-parametri optimizați poate estima cu precizie indicele de degradare pe un set de date de testare separat, obținând astfel un scor R^2 de 1 și o eroare absolută medie de aproximativ 0,1. Considerăm că aceste rezultate sunt foarte promițătoare în materie de performanțe, având în vedere setul extins de date utilizat, de numărul minim de caracteristici și de simplitatea algoritmului de învățare automată propus.

Trebuie menționat că, utilizarea unui set de date atât de mare și divers pentru o aplicație de aceste tip, atenuează riscul de overfitting a algoritmului nostru de învățare automată, contribuind astfel la robustețea modelelor de estimare propuse în acest raport.

Principala contribuție a autorilor este dezvoltarea unei aplicații numerice avansate care încorporează două caracteristici principale: o rutină de calcul numeric bazată pe Metoda Analogiei Fortei care este capabilă să efectueze simulări dinamice neliniare rapide și, un framework de analiză a performanțelor algoritmilor de învățare automate care poate fi utilizat într-un mod eficient pentru a determina cele mai bune caracteristici și cel mai bun algoritm de estimare a indicelui de degradare.

În plus, acest raport servește ca o resursă valoroasă pentru cercetători și practicieni care urmăresc să stabilească o

abordare simplificată pentru estimarea degradărilor structurale. Instrumentul numeric este accesibil la adresa web [18], oferind astfel o resursă practică pentru comunitatea științifică.

Cercetările viitoare vor lua în considerare dezvoltarea unor proceduri rapide de simulare neliniară pentru o gamă mai largă de sisteme structurale și utilizarea unor algoritmi mai avansați, cum ar fi Random Forrest, pentru o mai bună estimare a indicelui de degradare.

REFERINȚE

- [1] Vasile Calofir, Mircea Stefan Simoiu, and Ruben Munteanu. DIGITERRA - Feature selection JupyterNotebook. https://mybinder.org/v2/zenodo/10.5281/zenodo.10200267/?labpath=feature_selection.ipynb.
- [2] Vasile Calofir, Mircea Stefan Simoiu, and Ruben Munteanu. DIGITERRA - Estimation Models Comparison JupyterNotebook. https://mybinder.org/v2/zenodo/10.5281/zenodo.10200267/?labpath=compare_models.ipynb.
- [3] Ruben Munteanu, Vasile Calofir, Florin Mota, and George Nica. Large scale damage assessment framework for buildings in urban areas; the effect of active control implementation. 02 2023.
- [4] George Bogdan Nica, Munteanu Ruben Iacob, Vasile Calofir, and Mihail Iancovici. Modelling nonlinear behavior of 3d frames using the force analogy method. *Structures*, 35:1162–1174, 2022.
- [5] Oliver Richard de Lautour and Piotr Omenzetter. Prediction of seismic-induced structural damage using artificial neural networks. *Engineering Structures*, 31(2):600–606, 2009.
- [6] Musa Arslan, M. Ceylan, and T. Koyuncu. An ann approaches on estimating earthquake performances of existing rc buildings. *Neural Network World*, 22:443–458, 01 2012.
- [7] Konstantinos Morfidis and Konstantinos Kostinakis. Approaches to the rapid seismic damage prediction of r/c buildings using artificial neural networks. *Engineering Structures*, 165:120–141, 2018.
- [8] Konstantinos Morfidis and Konstantinos Kostinakis. Seismic parameters' combinations for the optimum prediction of the damage state of r/c buildings using neural networks. *Advances in Engineering Software*, 106:1–16, 2017.
- [9] Tiago Ferreira, João Estêvão, Rui Maio, and Romeu Vicente. The use of artificial neural networks to estimate seismic damage and derive vulnerability functions for traditional masonry. *Frontiers of Structural and Civil Engineering*, 14, 04 2020.
- [10] Kevin K. F. Wong and Rong Yang. Inelastic dynamic response of structures using force analogy method. *Journal of Engineering Mechanics*, 125(10):1190–1199, 1999.
- [11] Li Gang and K.F. Wong Kevin. Theory of nonlinear structural analysis: The force analogy method for earthquake engineering. pages 1–352, 03 2014.
- [12] G.C. Hart and K. Wong. *Structural Dynamics for Structural Engineers*. Wiley, 1999.
- [13] Munteanu Ruben Iacob, Florin Moța, Vasile Calofir, and Cătălin Baciu. New approach to nonlinear dynamic analysis of reinforced concrete 3d frames; an accurate and computational efficient mathematical model. *Applied Sciences*, 12(3), 2022.
- [14] George-Bogdan Nica, Florin Pavel, and Gabriel Hojda. A fast nonlinear dynamic analysis automated approach to produce fragility curves for 3d rc frames. *Engineering Structures*, 281:115695, 2023.
- [15] Zhongheng Zhang. Introduction to machine learning: k-nearest neighbors. *Annals of Translational Medicine*, 4(11):218–218, June 2016.
- [16] Anabella Cotovanu and Radu Vacareanu. Local site conditions modeling in stochastic simulation of ground motions generated by vrancea (romania) intermediate-depth seismic source. *Journal of Seismology*, 24:229–241, 02 2020.
- [17] Kenneth V. Price, Rainer M. Storn, and Jouni A. Lampinen. *Differential evolution: a practical approach to global optimization*. Natural computing series. Springer, Berlin ; New York, 2005.
- [18] Vasile Calofir, Mircea Stefan Simoiu, and Ruben Munteanu. DIGITERRA - Digital platform for estimating the building degradation index with machine learning techniques, November 2023. <https://zenodo.org/doi/10.5281/zenodo.10200267>.

Tabelul I: Setul de date ce descriu caracteristicile clădirilor regulate

Număr de etaje (no story) - no_story										
Nr. de etaje	4	5	6	7	8	9	10	11	12	13
Nr. de clădiri	2795	3225	3415	3155	3164	2887	2865	2983	2676	2832
Înălțime etaje - Hieght										
Intervale (m)	3.0-3.1	3.1-3.2	3.2-3.3	3.3-3.4	3.4-3.5	3.5-3.6	3.6-3.7	3.7-3.8	3.8-3.9	3.9-4.0
Nr. de clădiri	2502	2772	2822	3033	2833	3148	2967	2910	2602	4408
Înălțime clădiri - b_Hieght										
Intervale (m)	12-16	16-20	20-24	24-28	28-32	32-36	36-40	40-44	44-48	48-52
Nr. de clădiri	3240	3658	3708	3803	3255	3026	3554	2821	1862	1070
Lățime secțiune stâlpi - b_st										
Intervale (m)	0.3-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0	1.0-1.1	1.1-1.2	1.2-1.3	1.3-1.4
Nr. de clădiri	4289	3363	3452	3560	3234	2772	2294	3294	2021	1718
Înălțime secțiune stâlpi - h_st										
Intervale (m)	0.3-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0	1.0-1.1	1.1-1.2	1.2-1.3	1.3-1.4
Nr. de clădiri	4373	3337	3316	3774	3269	2594	2336	3112	2206	1680
Încărcarea uniform distribuită - py										
Intervale (kN/m)	6.0-6.8	6.8-7.6	7.6-8.4	8.4-9.2	9.2-10.0	10.0-10.8	10.8-11.6	11.6-12.4	12.4-13.2	13.2-14.0
Nr. de clădiri	2946	3138	3012	3046	2849	2903	3694	2083	3014	3312
Prima perioadă naturală de vibrație - T1										
Intervale (s)	0.4-0.6	0.6-0.8	0.8-1.0	1.0-1.2	1.2-1.4	1.4-1.5	1.5-1.7	1.7-1.9	1.9-2.1	2.1-2.3
Nr. de clădiri	2444	9220	7766	4909	2907	1566	767	304	77	37
A doua perioadă naturală de vibrație - T2										
Intervale (s)	0.3-0.5	0.5-0.7	0.7-0.9	0.9-1.1	1.1-1.3	1.3-1.5	1.5-1.7	1.7-1.9	1.9-2.1	2.1-2.3
Nr. de clădiri	1703	10155	8106	5046	2734	1397	498	254	87	17
A treia perioadă naturală de vibrație - T3										
Intervale (s)	0.3-0.4	0.4-0.6	0.6-0.7	0.7-0.9	0.9-1.0	1.0-1.2	1.2-1.3	1.3-1.5	1.5-1.6	1.6-1.8
Nr. de clădiri	957	7239	9378	6218	3485	1867	685	131	24	13
Numărul maxim de articulații plastice formate împărțite la numărul de articulații plastice definite - no_ph_max/nph										
Intervale	0.0-0.1	0.1-0.1	0.1-0.2	0.2-0.3	0.3-0.3	0.3-0.4	0.4-0.4	0.4-0.5	0.5-0.6	0.6-0.6
Nr. de clădiri	3294	2704	5086	5610	5471	4207	2336	999	253	37
Energia histeretică împărțită la energia intrării seismice - HE(i)/IE(i)										
Intervale	0.0-0.1	0.1-0.2	0.2-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.5	0.5-0.6	0.6-0.7	0.7-0.8
Nr. de clădiri	3625	1617	1888	2141	2733	3155	3823	4774	4789	1452
Modulul lui Young al betonului - E										
Intervale*1e+07 (kN/m ²)	1.3-1.5	1.5-1.8	1.8-2.0	2.0-2.3	2.3-2.5	2.5-2.7	2.7-2.9	2.9-3.2	3.2-3.5	3.5-3.7
Nr. de clădiri	27	31	364	4190	4343	4244	4253	4496	3987	4062
Forța tăietoare de bază maximă pe direcția X - FbXmax										
Intervale*1e+04 (kN)	0.04-0.3e	0.3-0.6	0.6-0.9	0.9-1.2	1.2-1.5	1.5-1.8	1.8-2.1	2.1-2.3	2.3-2.6	2.6-2.9
Nr. de clădiri	4749	9862	7441	4163	2204	973	385	126	67	27
Forța tăietoare de bază maximă pe direcția Y - FbYmax										
Intervale*1e+03 (kN)	0.4-3.5	3.5-6.6	6.6-9.7	9.7-13	13-16	16-19	19-22	22-25	25-28	28-31
Nr. de clădiri	6255	10634	7124	3356	1571	625	278	118	29	7
Numărul maxim de articulații plastice formate - nr_ph_max										
Intervale*1e+02	0-2.1	2.1-4.1	4.1-6.2	6.2-8.2	8.2-10	10-12	12-14	14-16	16-18	18-21
Capacitatea de incoavire a grinzilor - Mgr										
Intervale*1e+02 (kNm)	0.9-2.2	2.2-3.5	3.5-4.8	4.8-6.1	6.1-7.4	7.4-8.7	8.7-10	10-11	11-13	13-14
Nr. de clădiri	3022	5908	6259	5207	4429	2242	1679	653	493	105
Capacitatea de incoavire a stâlpilor pe direcția X - MstX										
Intervale*1e+03(kNm)	0.5-0.8	0.8-1.1	1.1-1.3	1.3-1.6	1.6-1.9	1.9-2.1	2.1-2.4	2.4-2.6	2.6-2.9	2.9-3.2
Nr. de clădiri	2986	4475	4499	4802	4050	3745	3031	1667	631	111
Capacitatea de incoavire a stâlpilor pe direcția Y - MstY										
Intervale*1e+03 (kNm)	0.6-0.8	0.8-1.1	1.1-1.3	1.3-1.5	1.5-1.7	1.7-1.9	1.9-2.1	2.1-2.4	2.4-2.6	2.6-2.8
Nr. de clădiri	3822	3283	4012	3993	3454	3247	3151	2782	1708	545
Număr de deschideri (no span) - no_span					Număr de travee (no bay) - no_bay					
Intervale	2	3	4	5	2	3	4	5		
Nr. de clădiri	7506	7377	7466	7648	7763	7509	7413	7312		
Înălțimea secțiunii grinzilor - h_gr										
Intervale (m)	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0			
Nr. de clădiri	201	1924	4277	6434	8527	7763	871			
Lățimea secțiunii grinzilor - b_gr										
Intervale (m)	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6						
Nr. de clădiri	9176	14975	5249	597						
Indicele de degradare - DI										
Intervale	0.0-0.2	0.2-0.5	0.5-0.7	0.7-1.0	1.0-1.2	1.2-1.5	1.5-1.7	1.7-2.0	2.0-2.2	2.2-2.5
Nr. de clădiri	9251	7419	4908	3031	1958	1278	924	574	421	233

Tabelul II: Setul de date ce descriu caracteristicile clădirilor neregulate

Număr de etaje (no story21) - no_story										
Number of story	4	5	6	7	8	9	10	11	12	13
Nr. de clădiri	2605	3285	3436	3213	2908	2749	2895	2813	2886	2782
Număr de etaje (no story22) - no_story2										
Number of story Intervale	2	3	4	5	6	7	8	9	10	11
Nr. de clădiri	8712	5864	4434	2967	2684	1611	1376	884	717	323
Înălțime etaje - Hieght										
Intervale (m)	3.0-3.1	3.1-3.2	3.2-3.3	3.3-3.4	3.4-3.5	3.5-3.6	3.6-3.7	3.7-3.8	3.8-3.9	3.9-4.0
Nr. de clădiri	2543	2572	2543	2757	2918	2920	2968	2865	3153	4333
Înălțime clădiri - b_Hieght										
Intervale (m)	12-16	16-20	20-24	24-28	28-32	32-36	36-40	40-44	44-48	48-52
Nr. de clădiri	3160	3682	3613	3524	3223	3079	3196	3020	1939	1136
Lățime secțiune stâlpi - b_st										
Intervale (m)	0.3-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0	1.0-1.1	1.1-1.2	1.2-1.3	1.3-1.4
Nr. de clădiri	4196	3246	3508	3568	3119	2387	2447	3255	2163	1683
Înălțime secțiune stâlpi - h_st										
Intervale (m)	0.3-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0	1.0-1.1	1.1-1.2	1.2-1.3	1.3-1.4
Nr. de clădiri	4230	3341	3319	3792	2856	2415	2669	3157	2059	1734
Încărcarea uniform distribuită - py										
Intervale (kN)	6.0-6.8	6.8-7.6	7.6-8.4	8.4-9.2	9.2-10.0	10.0-10.8	10.8-11.6	11.6-12.4	12.4-13.2	13.2-14.0
Nr. de clădiri	2892	2701	2793	2971	2778	3084	3877	2128	2775	3573
Prima perioadă naturală de vibrație - T1										
Intervale (s)	0.3-0.5	0.5-0.8	0.8-1.0	1.0-1.2	1.2-1.4	1.4-1.7	1.7-1.9	1.9-2.1	2.1-2.3	2.3-2.5
Nr. de clădiri	2641	10472	7586	4970	2409	1081	310	92	10	1
A doua perioadă naturală de vibrație - T2										
Intervale (s)	0.3-0.5	0.5-0.7	0.7-0.8	0.8-1.0	1.0-1.2	1.2-1.4	1.4-1.6	1.6-1.7	1.7-1.9	1.9-2.1
Nr. de clădiri	2679	9504	7433	4968	2601	1481	596	245	58	7
A treia perioadă naturală de vibrație - T3										
Intervale (s)	0.2-0.4	0.4-0.5	0.5-0.6	0.6-0.8	0.8-0.9	0.9-1.1	1.1-1.2	1.2-1.3	1.3-1.5	1.5-1.6
Nr. de clădiri	1765	8159	8291	5700	3158	1569	599	237	87	7
Numărul maxim de articulații plastice formate împărțite la numărul de articulații plastice definite - nr_ph_max/nph										
Intervale	0.0-0.1	0.1-0.1	0.1-0.2	0.2-0.3	0.3-0.3	0.3-0.4	0.4-0.5	0.5-0.5	0.5-0.6	0.6-0.6
Nr. de clădiri	3812	3688	4735	5085	4841	3772	2306	1078	229	26
Energia histeretică împărțită la energia intrării seismice - HE(i)/IE(i)										
Intervale	-0.0-0.1	0.1-0.2	0.2-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.6	0.6-0.7	0.7-0.8
Nr. de clădiri	3408	2043	2070	2481	2928	3479	3991	4317	3790	1065
Modulul lui Young al betonului - E										
Intervale*1e+07 (kN/m ²)	1.3-1.6	1.6-1.8	1.8-2.0	2.0-2.3	2.3-2.5	2.5-2.8	2.8-3.0	3.0-3.2	3.2-3.5	3.5-3.7
Nr. de clădiri	25	26	758	4388	4206	3963	4244	4102	4049	3811
Forța tăietoare de bază maximă pe direcția X - FbXmax										
Intervale*1e+03 (kN)	0.4-4.9	4.9-9.5	9.5-14	14-18	18-23	23-27	27-32	32-36	36-41	41-45
Nr. de clădiri	6652	11269	6807	3068	1210	377	126	37	19	7
Forța tăietoare de bază maximă pe direcția Y - FbYmax										
Intervale*1e+03 (kN)	0.6-5.3	5.3-9.9	9.9-15	15-19	19-24	24-28	28-33	33-38	38-42	42-47
Nr. de clădiri	6460	11342	6806	3087	1225	448	128	33	27	16
Numărul maxim de articulații plastice formate - nr_ph_max										
Intervale*1e+02	0-2.7	2.7-5.3	5.3-8.0	8.0-11	11-13	13-16	16-19	19-21	21-24	24-27
Nr. de clădiri	13611	8613	4120	1928	823	274	126	61	9	7
Capacitatea de înconvoiere a grinzilor - Mgr										
Intervale*1e+02 (kNm)	0.9-2.2	2.2-3.5	3.5-4.8	4.8-6.1	6.1-7.4	7.4-8.7	8.7-10	10-11	11-13	13-14
Nr. de clădiri	2893	6083	6619	5223	3914	2141	1597	557	439	106
Capacitatea de înconvoiere a stâlpilor pe direcția X - MstX										
Intervale*1e+03 (kNm)	0.5-0.8	0.8-1.1	1.1-1.3	1.3-1.6	1.6-1.9	1.9-2.1	2.1-2.4	2.4-2.6	2.6-2.9	2.9-3.2
* Nr. de clădiri	3023	4206	4650	4580	3946	3745	2911	1718	691	102
Capacitatea de înconvoiere a stâlpilor pe direcția Y - MstY										
Intervale*1e+03 (kNm)	0.6-0.8	0.8-1.1	1.1-1.3	1.3-1.5	1.5-1.7	1.7-1.9	1.9-2.1	2.1-2.4	2.4-2.6	2.6-2.8
Nr. de clădiri	3810	3036	4110	3851	3391	2806	3208	3029	1750	581
Dimensiune travee - bay2										
Intervale (m)	5.0-5.1	5.1-5.2	5.2-5.3	5.3-5.4	5.4-5.5	5.5-5.6	5.6-5.7	5.7-5.8	5.8-5.9	5.9-6.0
Nr. de clădiri	2588	2776	2768	2593	2825	2891	2764	2819	2593	4955
Înălțime secțiune grinzi - h_gr										
Intervale (m)	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0			
Nr. de clădiri	124	1903	4514	6672	8188	7457	714			
Număr de deschideri (no span21) - no_span					Număr de deschideri (no span22) - no_span2					
Intervale	2	3	4	5	2	3	4	5		
Nr. de clădiri	7671	7454	7291	7156	7341	7626	7490	7115		
Număr de travee (no bay21) - no_bay					Număr de travee (no bay22) - no_bay2					
Intervale	1	2	3	4	1	2	3	4		
Nr. de clădiri	15555	7943	4362	1712	6009	6063	6113	11387		
Lățime secțiune grinzi - b_gr										
Intervale (m)	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6						
Nr. de clădiri	9419	14853	4764	536						
Damage Index - DI										
Intervale	0.0-0.2	0.2-0.5	0.5-0.7	0.7-1.0	1.0-1.2	1.2-1.5	1.5-1.7	1.7-2.0	2.0-2.2	2.2-2.5
Nr. de clădiri	9992	8101	4812	2598	1562	950	617	453	269	218

Tabelul III: Setul de date cu principalele caracteristici ale inputurilor seismice

Amplitudinea maximă a accelerației seismice pe direcția X										
Intervale (m/s^2)	0.7-1.7	1.7-2.7	2.7-3.8	3.8-4.8	4.8-5.8	5.8-6.8	6.8-7.8	7.8-8.8	8.8-9.8	9.8-10.8
Nr. de accelerograme	9868	16189	15326	7985	4953	2750	1418	812	253	15
Amplitudinea maximă a accelerației seismice pe direcția Y										
Intervale (m/s^2)	0.7-1.8	1.8-2.9	2.9-4.1	4.1-5.2	5.2-6.3	6.3-7.4	7.4-8.5	8.5-9.6	9.6-10.8	10.8-11.9
Nr. de accelerograme	11720	17104	13399	6780	4873	2546	1604	1179	336	28
Frecvență fundametală a înregistrărilor seismice - Fund Freq1 & Fund Freq2										
Intervale (Hz)	0.4-0.7	0.7-1.1	1.1-1.4	1.4-1.8	1.8-2.1	2.1-2.4	2.4-2.8	2.8-3.1	3.1-3.4	3.4-3.8
Nr. de accelerograme	12	2	1	0	0	0	0	2	2	1