

UNIVERSITATEA DE MEDICINĂ, FARMACIE, ȘTIINȚE ȘI TEHNOLOGIE
“GEORGE EMIL PALADE” DIN TÂRGU MUREȘ
ȘCOALA DOCTORALĂ DE LITERE, ȘTIINȚE UMANISTE ȘI APLICATE
DOMENIU: INFORMATICĂ

REZUMAT TEZĂ DOCTORAT

**MODELAREA SISTEMELOR NELINIARE FOLOSIND TEHNICI
DE ÎNVĂȚARE AUTOMATĂ CU APLICAȚII ÎN DETECȚIA DE
ANOMALII**

Doctorand:

Roland BOLBOACĂ

Conducător Științific:

Prof. Ing. Habil. Dr. Béla GENGE

TÂRGU MUREȘ

2024

Lista de Publicații

Evaluarea publicațiilor respectă standardele stabilite de CNATDCU (Consiliul Național de Atestare a Titlurilor, Diplomelor și Certificatelor Universitare), aplicabile studenților doctoranzi înscriși după 1 octombrie 2018. Puntajele sunt prezentate pe baza clasificării conferințelor¹ și a revistelor² în domeniul Informatică.

1. **Bolboacă Roland**, Haller Piroska, & Genge Béla (2024, -). *Feature Analysis and Ensemble-based Fault Detection Techniques for Nonlinear Systems* Neural Computing and Applications. [Decision: Major Revision]
Articol Jurnal, Categorie: B, 0 puncte, sursă: UEFISCDI 2023².
2. **Bolboacă Roland**, Haller Piroska, & Genge Béla (2024, -). *Evaluation Techniques for Long Short-Term Memory Models: Overfitting Analysis and Handling Missing Values*. In the 37th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems. [Accepted]
Articol Conferință, Categorie: C, 2 puncte, sursă: CORE2023¹.
3. **Bolboacă Roland**, & Genge Béla (2023, October). *Unsupervised Outlier Detection in Continuous Nonlinear Systems: Hybrid Approaches with Autoencoders and One-Class SVMs*. In International Conference Interdisciplinarity in Engineering. Cham: Springer International Publishing.
Articol Conferință, Categorie: Not Ranked, 1 puncte, sursă: CORE¹.
4. **Bolboacă Roland**, & Haller Piroska (2023, March). *Performance Analysis of Long Short-Term Memory Predictive Neural Networks on Time Series Data*. Mathematics, 11(6), 1432.
Articol Jurnal, Categorie: C, 2 puncte, sursă: UEFISCDI 2022 (Octombrie)²
5. **Bolboacă Roland** (2022, October). *Adaptive ensemble methods for tampering detection in automotive aftertreatment systems*. IEEE Access, 10, 105497-105517.
Articol Jurnal, Categorie: B, 4 puncte, sursă: UEFISCDI 2021².
6. **Bolboacă Roland**, Haller Piroska, Kontses Dimitris, Papageorgiou-Koutoulas Alexandros, Doulgeris Stylianos, Zingopis Nikolaos, & Samaras Zissis (2022, June). *Tampering detection for automotive exhaust aftertreatment systems using long*

short-term memory predictive networks. In 2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW) (pp. 358-367). IEEE.

Articol Conferință, Categorie: NEW, 0 puncte, sursă: CORE2021¹.

7. **Bolboacă Roland**, Lenard Teri, Genge Béla, & Haller Piroška (2020, August). *Locality sensitive hashing for tampering detection in automotive systems*. In Proceedings of the 15th International Conference on Availability, Reliability and Security (pp. 1-7). 9th International Workshop on Cyber Crime.

Articol Conferință, Categorie C, 1 puncte, sursă: CORE2020¹.

8. Lenard Teri, **Bolboacă Roland**, Genge Béla, & Haller Piroška (2020, June). *MixCAN: Mixed and backward-compatible data authentication scheme for controller area networks*. In 2020 IFIP Networking Conference (Networking) (pp. 395-403). IEEE.

Articol Conferință, Categorie A, 4 puncte, source: CORE2020¹.

9. Lenard Teri, **Bolboacă Roland**, & Genge Béla (2020, September). *LOKI: A lightweight cryptographic key distribution protocol for controller area networks*. In 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 513-519). IEEE.

Articol Conferință, Categorie: National, 1 punct, sursă: CORE2020¹.

10. **Bolboacă Roland**, Genge Béla, & Haller Piroška (2019, September). *Using Side-Channels to Detect Abnormal Behavior in Industrial Control Systems*. In 2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 435-441). IEEE.

Articol Conferință, Categorie C, 2 puncte, sursă: CORE2018¹.

Distribuția punctajelor pe categorii:

- Categorie A: 4 puncte
- Categorie B: 4 puncte
- Categorie C: 7 puncte
- Categorie D: 2 puncte

Punctaj final: 17 puncte.

¹CORE Conference Ranking Portal. <http://portal.core.edu.au/conf-ranks/>

²UEFISCDI. <https://uefiscdi.gov.ro/premierea-rezultatelor-cercetarii-articole>

Chapter 1

Introducere

Sistemele neliniare sunt utilizate într-o gamă diversă de domenii. După cum este descris de Schoukens și Ljung în [1], orice sistem care se abate de la liniaritate este considerat neliniar. Într-un context general, sistemele neliniare sunt caracterizate prin absența unei relații liniare între intrările și ieșirile lor.

Observațiile generate de aceste sisteme sunt de obicei sub forma de seriilor de timp. Seriile de timp reprezintă o colecție continuă de valori măsurate secvențial și pot provenii din aproape toate domeniile științifice, în care măsurătorile sistemului sunt colectate sau măsurate în timp. În majoritatea cazurilor, aceste date provin din observații în timp real sau măsurători ale senzorilor sistemului.

Un model poate fi definit ca o reprezentare simplificată a relației dintre intrările și ieșirile sistemului, sau așa cum este descris de Bala et al. [2] ca "Un substitut pentru orice obiect sau sistem. O descriere scrisă a unui sistem este un model care prezintă un aspect al realității. Modelul de simulare este complet și descrie comportamentul dinamic al sistemului". Din perspectiva strictă a modelării sistemelor neliniare, se utilizează trei abordări principale, respectiv modelarea white-box, black-box și grey-box.

Învățarea automată devine o tehnică importantă pentru modelarea black-box, pe măsură ce complexitatea datelor și nevoia de predicții precise au crescut. O clasă specifică de algoritmi de învățare automată, în special rețelele neuronale, este utilizată pentru modelarea neliniară [3]. Printre modelele populare utilizate pentru modelarea sistemelor neliniare, Rețelele Neuronale Recurente (RNN), Long Short-Term Memory (LSTM) și Rețelele Neuronale Autoregresive Neliniare cu Intrări Externe (NARXNN) au câștigat popularitate datorită capacității lor de a captura dependențele temporale și de a se adapta la scenarii dinamice din lumea reală [4]. Cu toate acestea, RNN-urile standard suferă de ceea ce este cunoscut sub numele de vanishing gradient și exploding gradient [5]. Mai mult, datorită arhitecturilor lor simple și problemelor menționate mai sus, RNN-urile standard nu pot învăța eficient secvențe lungi [6]. Pentru a aborda aceste provocări, au fost dezvoltate modele avansate, cum ar fi LSTMurile. Aceste modele

avansate incorporează mecanisme de porți împreună cu concepte precum memorie pe termen scurt și lung.

În scenariile din viața reală, aceste sisteme pot fi distribuite pe zone geografice mari, cu senzori dispersați care furnizează citirile. Aceasta ridică întrebarea dacă astfel de sisteme ar trebui modelate folosind un singur model sau utilizând mai multe modele reduse pentru fiecare componentă sau subcomponentă a sistemului. Abordarea aici ar fi să utilizăm un ansamblu [7].

În general, orice ansamblu poate fi văzut și definit folosind trei caracteristici care afectează performanța sa. Prima este dependența de modelele de bază antrenate, fie secvențial, fie paralel. A doua caracteristică este metoda de fuziune, care implică alegerea unui proces potrivit pentru combinarea ieșirilor modelelor de bază folosind diferite abordări de vot ponderat sau metode de învățare meta. A treia caracteristică este heterogenitatea modelelor de bază, fie omogene, fie eterogene. Deși metodele bazate pe ansamble datează de peste două decenii, lucrările recente încă dovedesc eficiența lor până în ziua de azi [8]. Similar, în domeniul detecției de anomalii, eficiența abordărilor poate fi îmbunătățită semnificativ prin utilizarea ansamblelor [9].

Trecând la aplicarea practică a tezei, în special în domeniul detecției de anomalii. După cum este ilustrat de Aggarwal et al. [10], trei metodologii primare sunt utilizate pentru detectarea anomaliilor: supervizate, nesupervizate și semisupervizate. În general, abordările supervizate oferă rezultate excelente, dar sunt limitate la detectarea modelelor anormale învățate. În plus, obținerea datelor etichetate pentru toate scenariile posibile ar putea fi dificilă sau nerealistă în anumite situații. Prin urmare, tehnicile de detecție a anomaliilor nesupervizate reprezintă un subiect interesant de cercetare.

În serii temporale continue și fluxuri de date, detectarea abaterilor de la comportamentul normal necesită utilizarea modelelor de predicție [11]. În astfel de cazuri, comportamentul normal poate fi modelat prin diverse tehnici, categorizând anomalii ca fiind devieri evidente sau subtile de la normalitate [12].

Ținând cont de definițiile anterioare, devine evident că detecția anomaliei în datele seriale continue care provin din sisteme neliniare prezintă numeroase provocări. În primul rând, modelarea seriei temporale sau a procesului generativ al seriei temporale trebuie abordată cu grijă, deoarece modelele nesigure și inexacte duc la predicții slabe, ceea ce se traduce la rândul său într-o detecție imprecisă. În al doilea rând, trebuie selectată o măsură adecvată de similitudine. În cele din urmă, tehnica de detecție trebuie proiectată având în vedere componenta temporală, natura ieșirii modelului și tipurile de devieri care se doresc a fi detectate.

Având o reprezentare precisă a sistemului neliniar folosind modele de predicție cu complexitate redusă, următorul pas implică proiectarea detectorilor eficienți. Acești detectori trebuie să utilizeze ieșirea predictorilor pentru a detecta devierile de la comportamentul normal învățat. Putem argumenta că astfel de detectoare pot fi proiectate pentru a identifica devierile cumulative folosind abordări punctuale sau

tehnici bazate pe ferestre glisante. Aceste abordări sunt, de asemenea, identificate de Bl'azquez-Garc'ia et al. într-o lucrare recentă în direcția detecției de anomalii [13].

Într-o direcție diferită, autorii din [13] subliniază faptul că majoritatea tehnicilor de detecție a anomaliilor analizate se concentrează pe detectarea părților seriei temporale care diferă semnificativ de valoarea așteptată. Cu toate acestea, unele puncte anormale, sau serii de puncte, ar putea să semene cu seria temporală fără anomalii, iar o metodologie de cumulare a micilor devieri în timp ar putea oferi rezultate mai bune în astfel de scenarii. Mai mult, o analiză detaliată a costurilor computaționale ale rulării unor astfel de tehnici în timp real este încă necesară.

1.1 Obiective de Cercetare

Deși accentul principal al tezei este pe proiectarea tehnicilor de detecție nesupervizate bazate pe predicție pentru serii de timp provenite din sisteme neliniare, au fost definite obiective specifice de cercetare (**RO**).

- **RO1:** Dezvoltarea unui model îmbunătățit de predicție utilizând modelul LSTM pentru serii de timp provenite din sisteme neliniare.
- **RO2:** Dezvoltarea unei metodologii de selecție a caracteristicilor folosind o abordare bazată pe analiza sensibilității. În plus, efectuarea unei analize detaliată a modelului de predicție în ceea ce privește gestionarea valorilor lipsă și a supra-antrenării.
- **RO3:** Elaborarea unei metodologii eficiente de detecție nesupervizată care utilizează ieșirea modelelor de predicție, în contextul în care predictorii sunt antrenați doar cu observații fără anomalii.
- **RO4:** Elaborarea unui nou ansamblu adaptiv de detectoare care utilizează o metodologie eficientă de agregare a deciziilor.
- **RO5:** Implementarea soluțiilor propuse în medii diverse (de exemplu, Python, MATLAB) și în sisteme cu capabilități computaționale limitate.
- **RO6:** Testarea și validarea metodelor propuse folosind seturi de date reale și sintetice provenite din sisteme neliniare. În plus, aplicarea tehnicilor de predicție și detecție propuse pentru a rezolva probleme din lumea reală.

1.2 Structura tezei

- **Capitolul 2:** Capitolul al doilea introduce și definește conceptele importante utilizate în întreaga teză, inclusiv contextele istorice. Acest capitol prezintă, de asemenea, studii relevante și recente legate de abordările prezentate în teză, inclusiv modelarea neliniară, abordările de învățare automată, tehnicile de optimizare a hiperparametrilor și soluțiile de detectare a anomaliilor.

- **Capitolul 3:** Al treilea capitol introduce modelul LSTMF pentru modelarea sistemelor neliniare. LSTMF cuprinde un model LSTM îmbunătățit cu o versiune modificată a algoritmului Teacher Forcing. În plus, acest capitol prezintă o analiză extensivă a hiperparametrilor și o comparație a performanței de predicție cu diverse modele recente din literatura de specialitate.
Compararea performanței de predicție este realizată între modelul propus LSTMF și diverse alte modele cum ar fi NARXNN, Perceptroni Multi Strat, Rețele Recurente, Support Vector Machines, modele ARX, și Random Forests.
- **Capitolul 4:** Al patrulea capitol introduce două metodologii de selecție a caracteristicilor care sunt utilizate pentru a reduce complexitatea modelului LSTMF. În plus, acest capitol abordează alte probleme semnificative, inclusiv supraadaptarea modelului și mecanismele pentru a gestiona datele lipsă.
- **Capitolul 5:** Al cincilea capitol ilustrează două direcții în domeniul detectării anomaliilor, și anume detectarea tentativelor de fraudare și a stărilor de avarie, împreună cu două abordări de detectare bazate pe ansamblu. Performanța ansamblelor de detectare a anomaliilor introduse în acest capitol este comparată cu numeroase abordări supravegheate și nesupravegheate.
- **Capitolul 6:** Al șaselea și ultimul capitol al tezei prezintă principalele concluzii, contribuțiile originale, posibilele direcții viitoare, finanțarea cercetării și participarea autorului în diverse proiecte de cercetare.

Chapter 2

Analiza Literaturii de Specialitate

2.1 Modelarea cu Învățare Automată și Predicție

Astăzi, algoritmi de învățare automată sunt utilizați pentru rezolvarea multor probleme din viața reală în diverse domenii, de la sănătate, finanțe și fabricație până la agricultură. Pe o scară mai largă, învățarea automată poate fi considerată un termen care cuprinde o gamă largă de algoritmi și modele propuse pentru sarcini specifice, inclusiv diagnosticarea medicală [14], întreținerea predictivă [15], atribuirea autorului textului [16], și detectarea anomaliilor [17]. În direcția modelării, tehnicile de învățare automată sunt utilizate pentru a crea modele precise bazate pe date. Aceste abordări sunt utilizate în diverse domenii, inclusiv științele pământului, diverse domenii industriale, matematică și fizică, și chiar modelarea acțiunilor utilizatorilor.

O subclasă specifică a învățării automate include rețelele neuronale. Încă de la prima mențiune a neuronilor artificiali acum aproape 80 de ani [18], rețelele neuronale au evoluat constant și au fost larg aplicate în numeroase domenii și sectoare. Pentru modelarea sistemelor neliniare, sunt utilizate diverse arhitecturi bine-cunoscute, inclusiv RNN-uri, LSTM-uri și NARXNN-uri [3].

În serii temporale continue și fluxuri continue de date, detectarea abaterilor de la comportamentul normal necesită utilizarea modelelor de predicție. Cu toate acestea, pentru ca rețelele neuronale să captureze eficient dependențele temporale și dinamica prezente în sistemele neliniare, care pot fi descrise de ecuații diferențiale, doar intrările actuale pot să nu fie suficiente. În astfel de cazuri, ieșirile sau stările anterioare ale sistemului modelat pot oferi informații importante despre comportamentul sistemului și pot fi necesare pentru predicții precise ale ieșirilor ulterioare [19].

În momentul modelării dinamicii sistemelor neliniare, există o abordare remarcabilă în care incorporarea ieșirii anterioare sau a stării sistemului este considerată ca o intrare suplimentară, îmbunătățind astfel capacitatea rețelei neuronale de a captura eficient comportamentul temporal și dinamica sistemului [20]. Acest lucru poate fi realizat

prin aplicarea unei tehnici cunoscute sub numele de Teacher Forcing, așa cum a fost introdusă de Williams și Zipser în [21].

Pe scurt, Teacher Forcing (TF) denotă un algoritm de antrenare pentru RNN-uri, în care în timpul antrenării valoarea reală a ieșirii (de exemplu, observată) este furnizată ca o intrare suplimentară, în timp ce în timpul inferenței modelul preia predicția din pasul anterior ca o intrare suplimentară.

În cartea lui Goodfellow [5], TF este prezentată și ca o tehnică de antrenare a rețelelor neuronale, aplicabilă rețelelor neuronale recurente. Autorul mai menționează că această tehnică provine din criteriul de probabilitate maximă, în care în faza de antrenare, rețeaua neuronală primește valoarea reală a ieșirii ca intrare în pasul următor.

2.2 Selectarea Hiperparametrilor, Compararea Modelelor și Analiza Caracteristicilor

Thomas Breuel, din echipa de cercetare Google, a studiat comportamentul și performanța clasificatoarelor de tipul LSTM în [22]. Acest studiu analizează comportamentul LSTM-urilor pentru diferiți hiperparametri, dar și cum alegerea nelinearităților afectează performanța. Printre hiperparametrii testați se numără rata de învățare, numărul de unități ascunse și dimensiunea mini-lotului. Autorii s-au concentrat pe clasificarea cifrelor pe două seturi de date de referință populare, respectiv MNIST, care este un set de date de clasificare a scrierii cifrelor, și UW3, care este o bază de date de evaluare OCR. Rezultatele au arătat că performanța clasificatoarelor LSTM depinde în principal de ratele de învățare, în timp ce dimensiunea loturilor au o influență mai mică. Mai mult, LSTMul fără conexiuni de vizualizare a obținut performanțe superioare.

Într-un studiu mai recent, Siarni-Namini et al. [23] au analizat performanța prognozei seriilor temporale folosind LSTM-uri unidirecționale și LSTM-uri bidirecționale (BiLSTMs). Autorii au comparat performanța ARIMA, a LSTM-urilor și a BiLSTM-urilor în contextul predicției datelor financiare. Un aspect interesant al acestei cercetări este performanța predicției atunci când datele sunt învățate în ambele direcții (adică de la trecut la viitor și de la viitor la trecut). Rezultatele lor au arătat că timpul de antrenare al BiLSTM-urilor a fost mai lent, dar a depășit modelele LSTM unidirecționale și ARIMA în ceea ce privește acuratețea predicției. Cu toate acestea, autorii nu au furnizat informații arhitecturale sau despre hiperparametrii testați ai rețelelor neuronale, sau dacă cele două arhitecturi au fost antrenate și testate cu același set de hiperparametri.

2.3 Detectarea Anomaliilor în Serii de Timp

Anomaliile sau valorile extreme, așa cum sunt definite de Mehrotra et al. în [24], sunt "devieri semnificative de la normă". Aici, norma definește normalitatea, fie că

este vorba despre o stare, un comportament sau anumite proprietăți ale sistemului, obiectului, procesului observat, sau chiar a unei persoane sau a unui grup de persoane. Această normalitate este definită în funcție de contextul și domeniul în care sunt aplicate abordările de detectare a anomaliilor. În timp ce în [24] termenii anomalie și valoare extremă sunt utilizați interschimbabil, autorii subliniază că în unele articole termenul anomalie este utilizat când se discută despre procese, iar valori extreme sunt introduse când se discută despre date.

Prima direcție din această teză se concentrează pe detectarea stărilor de avarie. Detectarea acestor stări este o aplicație specifică a detectării anomaliilor care se concentrează pe identificarea devierilor de la comportamentul normal, cauzate fie de defecte ale senzorilor, defecte ale componentelor sau intervenții asupra sistemului. Yoon și Macgregor în [25] afirmă că "Scopul detectării defectelor este de a determina producerea unui eveniment anormal într-un proces". Luând în considerare sistemele mari, uneori distribuite, efectele unor defecte ar putea propaga între subsistemele sau componentele interconectate ale sistemului mai mare. Acest efect este vizibil în seriile temporale generate de sistem.

A doua direcție tratată în această teză se concentrează pe detectarea anomaliilor care sunt explicit ascunse de către o persoană malefică. Aceasta denotă un procedeu care modifică comportamentul sistemului pentru a obține anumite avantaje (de exemplu, financiare, operaționale). Mai mult, pentru ca intervenția să rămână nedescoperită, aceeași persoană malefică ascunde efectele modificărilor prin injectarea de citiri false care imită comportamentul normal al sistemului. Aceste citiri false injectate urmează o distribuție similară cu cea a condițiilor normale de funcționare a sistemului, și de asemenea ascund citirile reale care ar reflecta efectele intervenției în anumite componente ale sistemului.

Chapter 3

LSTMTF: O variantă Îmbunătățită LSTM pentru Modelarea Sistemelor Neliniare

Versiunea originală TF, așa cum a fost propusă pentru modelele RNN standard, folosește valoarea de ieșire observată anterior ca o intrare suplimentară în timpul antrenamentului, iar în timpul inferenței ieșirea modelului este introdusă din nou ca intrare. Acest capitol descrie un model LSTM îmbunătățit numit LSTMTF, care combină LSTM-ul standard cu TF într-un mod inovator. Modelul LSTMTF utilizează valoarea de ieșire observată anterior ca o intrare suplimentară în timpul atât al procedurii de antrenament, cât și al inferenței. Acest capitol descrie, de asemenea, modelul LSMTFC, care denotă un model LSTM standard cu algoritmul TF, în care valoarea de ieșire observată anterior este utilizată doar în timpul antrenării.

Efectele diferiților hiperparametri sunt analizate pentru noile modele introduse LSTMTF, LSMTFC și modelele LSTM standard, utilizând o abordare empirică. Hiperparametrii includ lungimea secvenței de intrare, numărul de întârzieri temporale, dimensiunea mini-lotului, rata de învățare și numărul de unități ascunse.

Modelele sunt antrenate și testate pe date provenind dintr-un sistem neliniar bine-cunoscut, respectiv din setul de date al procesului Tennessee Eastman [26, 27].

În continuare, performanțele de predicție ale diferitelor variante ale modelului LSTM sunt comparate cu modelul NARXNN. Autorii unor studii, precum cele din [23], au efectuat mii de experimente în această direcție. Similar, pentru acest capitol, aproximativ 100.000 de modele au fost antrenate și testate folosind o gamă largă de hiperparametri, configurații și moduri de predicție.

Acest capitol studiază, de asemenea, efectul de expunere [28] pentru rețelele neuronale antrenate cu TF în forma sa originală. Deși expunerea poate afecta performanța de predicție a rețelelor neuronale, acest capitol va discuta și avantajele obținute prin utilizarea LSTM-urilor cu TF pentru detectare anomaliilor.

În plus, pentru a valida în continuare modelul LSTMTF, sunt efectuate experimente suplimentare. Aici, modelul este comparat cu 11 modele de regresie și predicție parametrice și non-parametrice din literatură.

3.1 Arhitectura Propusă

3.1.1 Modelul Long Short-Term Memory

Modelul LSTM poate fi definit ca o versiune îmbunătățită a unei RNN, capabilă să captureze dependențele pe termen lung și scurt din secvențele de date, rezolvând în același timp problema gradientului care explodează și a gradientului care dispare [29]. Modelul standard LSTM va fi denumit în continuare LSTM Vanilla (VLSTM). Acest model poate cuprinde mai multe straturi care încorporează unități LSTM secvențiale. Aceste unități iau vectorul de intrare curent, denumit $X(t)$, împreună cu vectorul de stare ascuns anterior al stratului, denumit mai departe $h(t)$. Vectorul de stare a celulei, denumit $C(t)$, este actualizat folosind trei porți, respectiv porțiunea de intrare $I(t)$, porțiunea de uitare $f(t)$ și porțiunea de ieșire $o(t)$.

3.1.2 Modelul Long Short-Term Memory cu Teacher Forcing

Pe scurt, LSTM-ul standard reprezintă o funcție neliniară a intrărilor și stărilor ascunse anterioare. Ulterior, LSTMTF reprezintă o funcție neliniară a intrărilor anterioare, a stărilor ascunse și a ieșirilor observate.

După cum s-a menționat mai sus, aplicarea TF implică modificarea procedurilor de antrenament prin adăugarea, la fiecare pas de timp, a unei intrări suplimentare, care este valoarea anterioară reală $y(t - 1)$. Această valoare este ulterior propagată către toate porțile LSTM. Această intrare suplimentară este necesară atât în timpul antrenamentului, cât și în timpul inferenței. Dacă în faza de inferență valoarea reală a ieșirii nu este disponibilă, aceasta este înlocuită cu valoarea prevăzută anterior. TF este aplicabilă modelelor care au o conexiune recurentă de la ieșire înapoi în model și poate fi folosită ca alternativă la Backpropagation through time (BPTT). Cu toate acestea, TF poate fi totuși aplicată în combinație cu BPTT pentru antrenarea modelelor cu recurente [5].

După cum este descris în [5], antrenarea modelelor cu versiunea originală a TF poate duce la rezultate slabe de predicție, deoarece în timpul inferenței modelul ar putea fi expus la diferite date. Aceasta este cunoscută sub numele de expunere la bias. Expunerea la bias apare atunci când un model de învățare automată nu este expus la o gamă suficient de diversă de date în timpul antrenamentului. În esență, expunerea la bias apare atunci când distribuția datelor văzute de model în timpul antrenamentului nu reflectă exact distribuția datelor cu care se va confrunța în lumea reală.

3.2 Analiza Hiperparametrilor

Performanța și capacitățile predictive ale celor trei modele (adică VLSTM, LSTMTF și LSTMTFC) sunt testate folosind două configurații distincte și în două moduri de funcționare.

În primul rând, în ceea ce privește numărul de intrări și ieșiri, modelele sunt testate ca fiind de tip multi-intrare, ieșire unică (MISO) și multi-intrare, ieșire multiplă (MIMO). Configurația MISO implică predicția unei variabile folosind mai multe variabile de intrare. În mod similar, configurația MIMO implică predicția mai multor variabile folosind mai multe variabile de intrare.

În al doilea rând, în ceea ce privește modurile de predicție, sunt urmate două abordări, respectiv de la mulți la mulți (M2M) și de la mulți la unul (M2O). În cazul M2O, modelele vor lua ca intrare o secvență de X_n^φ intrări; aici, φ denotă numărul de pași de timp (cu alte cuvinte, lungimea secvenței de intrare) și n denotă numărul de variabile de intrare și va produce doar valoarea predictivă finală a secvenței. Pentru a exemplifica, pentru fiecare secvență de zece valori de intrare, modelul va produce următoarea valoare din secvență.

În cazul M2M, modelele vor lua similar ca intrare o secvență de X_n^φ valori și vor produce o altă secvență de valori de dimensiune φ , prima predicție începând la $t_0 + 1$. Aici, t_0 denotă momentul primei valori din secvența de intrare.

3.3 Rezumat

Acest capitol a prezentat LSTMTF, un model LSTM combinat cu o nouă variantă a algoritmului TF aplicată atât în timpul antrenamentului, cât și în inferență. De asemenea, acest capitol a oferit o analiză detaliată a modelelor LSTM antrenate cu și fără TF. TF a fost aplicat în două variante, cu valoarea actuală (observată) returnată ca intrare atât în timpul antrenamentului, cât și în testare, așa cum a fost propus pentru sarcinile de detectare a anomaliilor (de exemplu, LSTMTF), și așa cum a fost propus inițial, cu valorile prezise returnate în timpul testării (de exemplu, LSTMTFC). Mai mult, acest capitol a prezentat și măsurători ale timpului de antrenament și testare pentru arhitecturile testate.

Procedurile de antrenament și testare au fost realizate folosind o gamă largă atât de hiperparametri interni, cât și de hiperparametri externi, în timp ce rezultatele au fost analizate folosind diverse metrice de performanță. Modelele au fost testate în mai multe configurații, respectiv multi-intrare, ieșire unică și multi-intrare, ieșire multiplă, folosind două moduri de predicție: de la mulți la mulți și de la mulți la unul. Pentru reproducibilitate, toate configurațiile testate ale rețelelor neuronale, hiperparametrii și seturile de date au fost documentate pe parcursul acestui capitol.

În ambele configurații, MISO și MIMO, VLSTM (adică LSTM-ul standard fără forțare de învățare) a obținut rezultate mai bune în ceea ce privește timpul de convergență al antrenamentului pentru metoda de predicție M2M; cu toate acestea, în ceea ce privește MAE-ul de predicție, LSTMTF a obținut rezultate mai bune. Din cele trei rețele neuronale, LSTMTF a obținut cele mai proaste rezultate în ceea ce privește MAE-ul de testare folosind abordarea M2M.

În fiecare scenariu experimental, a existat o mică scădere a erorii de predicție atunci când s-a trecut de la configurația MISO la configurația MIMO pe același tip de rețea neuronală. Cu toate acestea, timpul necesar pentru antrenament și testare a crescut cu 21 % pentru antrenament și cu 28% pentru testare atunci când s-a folosit configurația MIMO. În general, lungimea secvenței de intrare, dimensiunea mini-lotului, numărul de unități ascunse și numărul de întârzieri au influențat performanța de antrenament și testare. În schimb, influența ratei de învățare a părut să fie mai mică în toate experimentele pentru toate arhitecturile.

După cum ilustrează rezultatele experimentale, arhitectura LSTMs poate fi redusă semnificativ, menținând în același timp performanța de predicție. Acest lucru s-a observat utilizând mai puține caracteristici pentru arhitectura MISO, obținând totuși rezultate similare cu arhitectura MIMO. În plus, utilizând TF, s-a arătat că modelele pot fi antrenate cu un număr redus de exemple, folosind doar un singur strat ascuns și totuși pot depăși alte modele. Arhitectura redusă (adică numărul de intrări, un singur strat ascuns și numărul de unități ascunse) și posibilitatea de a antrena modele cu mai puține exemple fac ca astfel de modele să fie candidați potriviți pentru operațiuni în timp real și pe dispozitive cu resurse limitate.

Modelul propus LSTMTF a fost de asemenea comparat cu 11 modele suplimentare de regresie și predicție parametrice și nonparametrice, inclusiv unele dintre cele mai recente. Aceste rezultate experimentale suplimentare ilustrează capacitățile de modelare ale LSTMTF-ului în ceea ce privește valorile MAE măsurate. În toate experimentele, modelul propus a obținut valori MAE mai mici în comparație cu cele 11 modele testate.

Chapter 4

Analiza de Caracteristici pentru LSTMTF

Acest capitol îmbunătățește modelul LSTMTF, așa cum a fost introdus în capitolul anterior, prin propunerea a două metodologii de selecție a caracteristicilor, o abordare de analiză a corelației și o metodă pentru gestionarea valorilor lipsă în timp real. De asemenea, acest capitol prezintă evaluări experimentale extinse și rezultate pentru toate metodele propuse.

4.1 Selectarea Caracteristicilor

În direcția selecției caracteristicilor, acest capitol introduce o metodă de selecție a caracteristicilor bazată pe corelație. Această metodă folosește scorul coeficientului lui Pearson pentru a selecta grupul de intrări pentru fiecare variabilă de ieșire. Ca parte a celei de-a doua abordări, este introdusă o nouă metodă inspirată de analiza sensibilității [30]. Această metodă include clasificarea caracteristicilor și selecția automată a caracteristicilor, pentru sarcinile de regresie. Deși această metodă este propusă și testată pe modelele LSTMTF, este potrivită și pentru alte modele. Procesul de selecție automată a caracteristicilor include o abordare proactivă, începând cu un număr prestabilit de caracteristici, și incorporând la fiecare pas caracteristici suplimentare pe baza importanței acestora, până când este îndeplinit un criteriu de oprire.

4.1.1 Selecție Bazată pe Corelație

Această abordare urmează o tehnică bazată pe corelație pentru selectarea grupurilor de intrări și ieșiri. Procesul de selecție folosește coeficientul de corelație Pearson pentru a descrie puterea relației dintre variabile. Din numeroasele semnale măsurate, cele care prezintă un coeficient de corelație ridicat cu semnalul de ieșire ales sunt selectate.

4.1.2 Selecție Bazată pe Scor

În rezumat, metodologia pentru clasificarea caracteristicilor utilizează o abordare retrogradă bazată pe analiza sensibilității, care nu necesită reantrenarea modelelor LSTM. Lista rezultată a caracteristicilor clasificate este folosită pentru selecția caracteristicilor. Procesul de selecție urmează o abordare proactivă în care, pe baza unui criteriu de oprire, caracteristicile cu cele mai mari clasificări sunt adăugate incremental, unul câte unul, iar modelul este reantrenat la fiecare pas. Aceleași proceduri de selecție a caracteristicilor sunt urmate pentru mai multe variabile de ieșire. În plus, în ambele abordări, distanța Cramer este utilizată pentru a calcula un scor de distanță.

4.2 Teste de Supra-antrenare

Pentru dezvoltarea testelor de supra-antrenare, se pornește de la premisa că modelele LSTM cu TF supra-antrenează valoarea de ieșire anterioară. Prin urmare, în timpul inferenței, aceste modele prezic aproape de valoarea de ieșire anterioară, ignorând relația spațio-temporală dintre restul intrărilor și variabila de ieșire.

În această direcție au fost create următoarele presupuneri. În primul rând, dezactivarea oricăreia dintre intrările suplimentare nu ar avea niciun efect asupra performanței modelului, deoarece acesta se "bazează" doar pe valoarea de ieșire anterioară pentru a face noi predicții. În al doilea rând, dezactivarea tuturor intrărilor suplimentare nu ar afecta performanța modelului, pe baza aceleiași presupuneri de mai sus. În al treilea rând, stabilirea valorii de ieșire anterioare la o constantă nu ar afecta performanța, deoarece modelul ar prezice doar aproape de valoarea anterioară de ieșire.

Abordarea propusă generează un scor de supra-antrenare, măsurând distanța între distribuția erorilor de predicție în diverse scenarii, folosind trei metrice de distanță a distribuției.

4.3 Valori Lipsă

Pentru modelul LSTMF, se presupune că valorile de ieșire anterioare sunt disponibile atât în timpul antrenamentului, cât și în timpul inferenței. Cu toate acestea, apare o problemă semnificativă în această abordare atunci când valorile de ieșire anterioare lipsesc din cauza evenimentelor neprevăzute (de exemplu, defecte de comunicare și citiri eronate ale senzorilor).

Deși există multe tehnici de imputare pentru a gestiona datele lipsă, foarte puține dintre ele se ocupă de gestionarea datelor lipsă în scenariile în timp real.

Ca soluție la această problemă, acest capitol introduce o abordare care nu necesită utilizarea unor modele suplimentare pentru imputare. Abordarea propusă pentru gestionarea valorilor lipsă implică comutarea aceluiași model de la utilizarea valorii de

ieșire anterioare, ca o intrare suplimentară, la utilizarea valorii sale prezise anterioare, în timp real, atunci când valorile lipsesc.

Pentru acest scenariu, se consideră cazul în care lipsesc doar câteva valori, datorită defectelor de comunicare sau a măsurătorilor eronate. Într-un scenariu real, valorile lipsă pe o perioadă îndelungată ar putea fi indicii clare ale unor defecțiuni (determinând inițierea mecanismelor interne de alertă) și ar putea duce chiar la oprirea mai multor sisteme, totuși, în cazul în care lipsesc doar câteva măsurători, soluția noastră propusă ar trebui să continue să funcționeze.

4.4 Rezumat

Acest capitol a prezentat două abordări pentru selecția caracteristicilor utilizând coeficientul de corelație al lui Pearson și o abordare inovatoare bazată pe scor. A doua abordare propusă este inspirată de analiza sensibilității și este aplicabilă în problemele de regresie. În această direcție, Distanța de Energie a fost utilizată pentru calcularea importanței caracteristicilor, împreună cu o metodologie de selecție a caracteristicilor. Rezultatele experimentale au ilustrat faptul că abordarea bazată pe scor poate obține rezultate remarcabile și comparabile cu tehnici bine stabilite, cu valori scăzute ale RMSE și valori ridicate ale R^2 .

În plus, acest capitol a propus o metodă inovatoare pentru testarea dacă modelele care utilizează TF depind puternic doar de valoarea anterioară a ieșirii și ignoră alte intrări exogene. În acest caz, aceste modele ar putea prezice în mod naiv valori noi incorecte. Această investigație s-a concentrat în mod specific pe modelul LSTMTF și a utilizat trei metrici de distanță a distribuției. Rezultatele evaluării experimentale au evidențiat faptul că toate metricile de distanță selectate au demonstrat că modelul nu s-a supra-antrenat în niciunul dintre scenariile testate. Acest capitol a introdus, de asemenea, o metodă pentru tratarea datelor lipsă în timp real.

Chapter 5

Detectarea Anomaliilor folosind LSTMF

Acest capitol explorează două direcții distincte din domeniul detectării anomaliilor în sistemele nonlineare, și anume detectarea fraudelor și detectarea defecțiunilor.

Un număr semnificativ de încercări de fraudă au fost observate în industria auto, vizând în special sistemele de protecție a mediului. Aceste încercări au fost identificate în numeroase studii, cum ar fi [31]. Mai mult, pe măsură ce gravitatea fraudelor a atins niveluri critice, proiecte ample de cercetare și inovație au fost dedicate eliminării fraudelor în sistemele relevante pentru emisii, de exemplu, proiectul Sisteme de Diagnostic Antifraudă (DIAS) [32].

Trecând la a doua direcție de aplicabilitate, respectiv detectarea stărilor de avarie (defecțiunilor). Detectarea acestora poate fi văzută ca o aplicație specifică a detectării anomaliilor, concentrată pe identificarea deviațiilor de la comportamentul normal al sistemului cauzate de defecțiuni ale senzorilor, defecțiuni ale componentelor sau intervenții în sistem. Yoon și Macgregor în [25] afirmă că "scopul detectării defecțiunilor este de a determina apariția unui eveniment anormal într-un proces". În plus, așa cum a fost evidențiat de Amini și Zhu [33], clasificarea greșită a eșantioanelor normale drept defecțiuni poate duce la perturbări operaționale inutile și costuri crescute.

5.1 Soluții Propuse de Detectare a Anomaliilor

Pentru a aborda direcțiile descrise anterior în domeniul detectării anomaliilor, respectiv detectarea fraudelor și detectarea defecțiunilor, acest capitol introduce două ansambluri de detectare a anomaliilor. Ansamblurile valorifică capacitățile de modelare ale LSTMF-urilor, care sunt utilizate ca mijloace de modelare a sistemelor nonlineare monitorizate. Ansamblurile incorporează, de asemenea, abordări de detectare bazate pe diagrama de control Cumulative Sum (CUSUM) și distanța histogramelor, care

monitorizează modificările în reziduurile prezicerilor LSTMF. Deciziile detectorilor sunt fuzionate folosind două tehnici bazate pe vot majoritar.

În faza de detectare, fiecare detector monitorizează un semnal (de exemplu, o variabilă) analizând deviațiile de la comportamentul normal învățat. Folosind o metodologie bazată pe prag, fiecare detector generează o decizie binară cu privire la validitatea unui nou punct de date. Decizia finală este dată de ansamblu folosind tehnici bazate pe vot majoritar.

5.1.1 Arhitectura Ansamblurilor

Un element comun în ansamble este modelul predictiv MISO al LSTMF-urilor. În ambele ansamble, detectoarele de bază monitorizează un semnal specific analizând constant erorile de predicție primite de la predictor, folosind două tehnici distincte. Aceste detectoare sunt Detectorul Bazat pe Sume Cumulative (CBD) și Detectorul Bazat pe Distanța Histogramelor (HBD).

Deciziile individuale ale detectorilor de bază sunt combinate folosind două metodologii. În primul rând, folosind un sistem de vot majoritar. În al doilea rând, folosind o metodologie adaptiv de fuziune a voturilor care consideră deciziile istorice ale fiecărui detector și generează una dintre cele trei decizii: normal, alertă și avertisment.

5.1.2 Arhitectura Detectorilor de Bază

CBD monitorizează modificările atât în valorile medii, cât și în cele ale varianței erorilor de predicție LSTMF, folosind două variante ale schemei 1-CUSUM. Schema 1-CUSUM are capacitatea de a detecta modificări (adică creșteri și scăderi) în valorile medii și de varianță, folosind o diagramă de control bidirecțională, care funcționează cu observații unice. Prima variantă propusă utilizează CUSUM așa cum a fost propusă inițial în [34], pentru o calculație CUSUM punct cu punct, în timp ce a doua variantă utilizează o metodologie cu fereastră glisantă, care calculează valorile CUSUM pe o fereastră glisantă.

Detectorul de bază al celui de-al doilea ansamblu, numit HBD, utilizează, de asemenea, modele predictive LSTMF, dar procesează erorile de predicție într-un mod diferit. În primul rând, HBD construiește histograma erorilor de predicție pe o fereastră de timp dată. În al doilea rând, folosind o metrică de distanță personalizată, calculează distanța între histograma erorilor de predicție a datelor de antrenament și histograma datelor conținute în fereastra de timp curentă. În cele din urmă, fiecare detector generează o decizie binară folosind o abordare bazată pe prag.

5.1.3 Agregarea Adaptivă Bazată pe Vot Majoritar

Tehnica de fuziune sau agregare propusă, aplicată pentru detectarea fraudelor, este o versiune modificată a metodei de votare ponderată a majorității, cu o componentă de

reputație istorică. Cu alte cuvinte, la fiecare pas, ponderile detectorilor sunt ajustate (de exemplu, crescute sau scăzute) dacă detectorul votează sau nu la fel ca majoritatea. Ponderile sunt ajustate printr-un procent, care este calculat pe baza deciziilor anterioare ale aceluși detector (de exemplu, dacă în trecut a votat la fel ca majoritatea). Această metodologie de ajustare acordă ponderi mai mari detectorilor care votează la fel ca majoritatea în comparație cu detectoarele care sunt mai des în dezacord cu majoritatea. În plus, ponderile pot deveni zero, astfel *ignorând temporar* decizia aceluși detector.

Decizia finală a ansamblelor poate fi una dintre următoarele: normal, alertă sau avertisment. Starea de avertisment este declanșată atunci când votul majoritar nu declanșează o alertă, dar există cel puțin un detector care a identificat o observație frauduloasă. Având în vedere că fiecare detector monitorizează un semnal diferit, există posibilitatea ca fraudarea unui sau mai multor semnale să nu afecteze restul semnalelor, în special dacă componenta fizică fraudată nu funcționează într-un circuit închis. Astfel, generând un avertisment, se poate efectua o investigație suplimentară asupra acelei componente.

5.2 Rezumat

Acest capitol s-a adresat în primul rând unei noi amenințări și anume fraudarea sistemelor de protecție a mediului ale vehiculelor. Frauda poate avea efecte grave atât asupra sănătății umane, cât și asupra mediului, deoarece vehiculele fraudate emit concentrații mai mari de poluanți, cum ar fi oxizii de azot. În plus, acest capitol a introdus, de asemenea, o abordare bazată pe ansamble pentru detectarea defecțiunilor în sistemele continue neliniare.

Acest capitol a introdus două metodologii bazate pe ansamble pentru detectarea fraudelor. Soluțiile propuse utilizează modele predictive LSTMFTF împreună cu detectoare bazate pe CUSUM și distanțe între histograme. Detectorii bazați pe CUSUM și histograme primesc ca intrare eroarea de predicție de la modelele predictive și generează o decizie binară folosind o abordare bazată pe prag.

Atunci când au fost aplicate pentru detectarea fraudelor, soluțiile propuse au obținut rezultate remarcabile, incluzând rate de fals pozitive de 0% pe toate seturile de date și până la 100% detecție. În plus, ansamblele au fost comparate cu metodologiile recente de detectare a fraudelor, cu rezultate promițătoare. Acest capitol oferă, de asemenea, măsurători ale consumului de resurse și scalabilității pe un sistem de referință, demonstrând posibilitatea integrării soluțiilor propuse într-un mediu real.

În contextul detectării defectelor, tehnica bazată pe ansamble a demonstrat eficiență remarcabilă prin obținerea unei rate de detectare de până la 100% cu o rată de fals pozitive de 0% pentru fiecare defecțiune necunoscută. Aceleași rezultate au fost observate chiar și în cazul defecțiunilor dificile. În plus, tehnica bazată pe ansamble a depășit numeroase alte metode din literatura științifică.

Chapter 6

Concluzii

În conformitate cu primul și al cincilea obiectiv de cercetare, al treilea capitol a introdus o abordare pentru modelarea sistemelor neliniare, și anume modelul LSTMTF. Acest model cuprinde un model LSTM cu o versiune modificată a algoritmului Teacher Forcing aplicat atât în timpul antrenării, cât și în inferență pentru seturile de date provenite din sisteme continue neliniare.

Pentru a valida modelul LSTMTF, s-a efectuat o analiză extensivă a hiperparametrilor pe un set de date de referință bine cunoscut. Această evaluare a inclus diferiți hiperparametri și mai multe arhitecturi. În plus, performanța de predicție a modelului LSTMTF a fost comparată cu 15 tehnici de modelare din literatură, cu rezultate promițătoare. Rezultatele experimentale au dezvăluit, de asemenea, că modelul LSTMTF poate modela comportamentul sistemelor neliniare chiar și cu arhitecturi reduse.

În conformitate cu al doilea și al cincilea obiectiv de cercetare, al patrulea capitol a introdus două metodologii de selecție a caracteristicilor. Prima abordare de selecție a caracteristicilor utilizează coeficientul de corelație al lui Pearson între posibili candidați de intrare și variabila de ieșire în timpul procesului de selecție. A doua abordare propusă cuprinde o metodologie nouă de selecție și calcul a importanței caracteristicilor. Această abordare utilizează mai întâi Distanța de Energie ca mijloc de a cuantifica distanțele dintre erorile de predicție și pentru a calcula un scor, care este ulterior utilizat în procedurile de selecție. A doua abordare de selecție a caracteristicilor a fost comparată experimental cu 7 abordări bine cunoscute din literatura.

După cum s-a menționat mai sus, modelul LSTMTF utilizează algoritmul TF în timpul atât al antrenării, cât și al inferenței, în care valoarea de ieșire observată anterior este introdusă ca o intrare suplimentară la fiecare pas de timp. Acest lucru a ridicat întrebarea dacă modelele care utilizează această abordare TF depind puternic doar de valoarea anterioară a ieșirii și ignoră celelalte intrări. În această direcție, au fost introduse și testate empiric trei presupuneri, utilizând o metodologie inspirată de analiza sensibilității. Pentru a analiza aceste presupuneri, au fost generate diferite scenarii prin

dezactivarea secvențială a intrărilor sau stabilirea unor valori constante pentru variabila de ieșire.

În plus, al patrulea capitol a introdus, de asemenea, o metodă pentru tratarea datelor lipsă în timp real. Rezultatele experimentale au evidențiat că utilizând această metodă, modelul poate gestiona eficient valorile lipsă îmbunătățind robustețea și fiabilitatea modelului LSTMF.

Pentru a aborda ultimele trei obiective de cercetare, în al cincilea capitol, modelul LSTMF a fost utilizat ca parte a două noi ansamble pentru detectarea anomaliilor. Aceste ansamble cuprind, de asemenea, detectoare bazate pe CUSUM și distanța histogramelor care sunt utilizate pentru a monitoriza erorile de predicție ale LSTMF-urilor. Detectorii bazati pe CUSUM funcționează în două moduri, punct-cu-punct și bazat pe fereastră glisantă. Pentru al doilea tip de detector, care monitorizează schimbările în distribuția erorilor, a fost propusă o nouă metrică de distanță care a fost demonstrată și formal. Deciziile detectorilor sunt agregate folosind două abordări. În primul rând, utilizând un sistem de vot majoritar. În al doilea rând, utilizând o nouă metoda adaptivă.

Ansambele propuse au fost validate în două scenarii distincte din domeniul detectării anomaliilor, respectiv fraudarea sistemelor auto și detectarea defectiunilor în sisteme neliniare. Rezultatele experimentale efectuate pe patru seturi de date au evidențiat validitatea metodelor propuse, dezvăluind rate scăzute de alerte false și rate ridicate de detecție în majoritatea scenariilor. În ambele scenarii ansamblele propuse au fost comparate cu 24 de soluții existente, cu rezultate promițătoare. În mod specific, ansamblurile propuse au depășit alte metode în ceea ce privește ratele de alerte false, ratele de detecție și întârzierile de detecție în majoritatea scenariilor experimentale.

Al cincilea capitol a prezentat, de asemenea, rezultatele măsurătorilor de resurse, efectuate pe un dispozitiv cu capacități limitate. Aceste măsurători au fost efectuate pentru a verifica posibilitatea integrării abordărilor noastre în dispozitive cu resurse limitate și au inclus măsurători CPU, memorie și scalabilitate. Rezultatele acestor experimente au demonstrat posibilitatea integrării soluțiilor propuse într-un mediu real, cu un impact redus asupra consumului de resurse.

6.1 Contribuții Științifice

Această secțiune rezumă contribuțiile științifice semnificative ale tezei.

- **Capitolul 2:**

1. Revizuire literaturii științifice în direcția tezei. Acest lucru include sisteme neliniare, modelare bazată pe învățare automată, arhitecturi de rețele neurale, selecție de hiperparametri, analiză de caracteristici și două domenii de aplicare pentru detectarea anomaliilor.

- **Capitolul 3:**

1. Proiectarea și dezvoltarea modelului LSTMTF, un model LSTM cu o variantă a metodei Teacher Forcing, pentru modelarea sistemelor neliniare folosind serii de timp.
2. Analiză extensivă a hiperparametrilor modelului LSTMTF pe un set de date popular provenind din reprezentarea unui sistem neliniar de referință. Această evaluare include diverși hiperparametri (de exemplu, lungimea secvenței, dimensiunea mini-lotului, rata de învățare, unitățile ascunse, întârzierile TF) și multiple arhitecturi de model și moduri de predicție.
3. Măsurători ale timpului de antrenare/testare pentru arhitecturile LSTMTF analizate.
4. Comparări extinse cu 15 abordări utilizate în literatura științifică pentru modelarea neliniară. Aceasta include modele de regresie și predicție parametrice și non-parametrice bine cunoscute.
5. Compararea performanței modelului LSTMTF și a altor algoritmi de predicție cu selecție manuală și optimizare automată a hiperparametrilor.

- **Capitolul 4:**

1. Proiectarea și dezvoltarea unei noi metodologii de analiză a caracteristicilor pentru modelele de regresie.
2. Compararea performanței cu 7 tehnici de selecție a caracteristicilor existente în literatura de specialitate.
3. Măsurători ale complexității și timpului de antrenare/testare pentru modelul LSTMTF cu și fără selecție a caracteristicilor.
4. Analiză exhaustivă a supra-antrenării modelului LSTMTF, utilizând trei metrici de distanță între distribuții și testând trei ipoteze.
5. Proiectarea unei metodologii pentru a trata valorile lipsă în timpul testării utilizând modelul LSTMTF. Valori lipsă cauzate de citiri eronate ale senzorilor sau erori de comunicare.

- **Capitolul 5:**

1. Proiectarea și dezvoltarea unui cadru de detectare a anomaliilor bazat pe ansamble care cuprinde următoarele componente:
 - Proiectarea unor detectori bazați pe sumă cumulativă și distanța dintre histograme, capabili de două moduri de funcționare, respectiv punct-cu-punct și bazat pe fereastră glisantă.
 - Proiectarea și dezvoltarea ansamblelor de detectare a anomaliilor nesupravegheate care înglobează predictorii LSTMTF împreună cu detectoarele proiectate a priori.
 - Proiectarea unei metodologii adaptive de agregare. Aceasta consideră deciziile precedente ale detectorilor în procedurile de ajustare a ponderilor.

2. Proiectarea și dezvoltarea unor noi abordări de selecție a parametrilor pentru ansamblurile de detectare a anomaliilor.
3. Aplicabilitatea ansamblurilor de detectare a anomaliilor pentru rezolvarea problemelor din lumea reală, în special manipularea autovehiculelor și detectarea defecțiunilor sistemelor neliniare, cu rezultate promițătoare.
4. În ambele direcții de detectare a anomaliilor, ansamblele sunt comparate cu un total de 24 abordări similare, cu rezultate promițătoare în ceea ce privește ratele ridicate de detecție, alertele false reduse și întârzierile de detectare. În plus, în direcția detectării defectelor: detectarea cu succes și cu o precizie ridicată a trei defecte considerate dificil de detectat de numeroși cercetători în literatura de specialitate.
5. Măsuri de performanță în ceea ce privește utilizarea CPU-ului, a memoriei și scalabilitatea pe dispozitive cu resurse limitate.

6.2 Direcții Viitoare

După cum a reieșit din al cincilea capitol, tehnicile supervizate care utilizează clasificatoare, antrenate cu observații atât anormale, cât și curate, obțin rezultate superioare. Cu toate acestea, raritatea seturilor de date disponibile necesită încă cercetări suplimentare în direcția metodelor nesuprvizate. O direcție interesantă pentru cercetările viitoare implică utilizarea abordărilor de detecție nesupravegheate capabile să identifice și să clasifice sursa anomaliei (de exemplu, semnalul anormal). Mai mult, integrarea tehnicilor de inteligență artificială explicabile rămâne o direcție deschisă și interesantă pentru cercetare.

O direcție suplimentară posibilă este explorarea diferitelor alte abordări și tehnici de detecție nesupravegheate, în afară de rețelele neurale. Cercetările viitoare ar putea să se concentreze pe dezvoltarea unor abordări care sunt antrenate cu un număr limitat de observații, dar sunt capabile să detecteze comportamente anormale cu o precizie ridicată. Direcțiile specifice de cercetare ar putea include introducerea de tehnici și algoritmi noi care pot adapta la comportamentele sistemelor evolutive.

Deși analiza caracteristicilor, împreună cu tehnicile de gestionare a datelor lipsă, rămân direcții extensiv cercetate, studii suplimentare în diverse alte domenii, pe seturi de date originare atât din sisteme liniare, cât și din cele neliniare, ar putea oferi soluții îmbunătățite.

Referințe

- [1] Johan Schoukens and Lennart Ljung. “Nonlinear system identification: A user-oriented road map”. In: *IEEE Control Systems Magazine* 39.6 (2019), pp. 28–99.
- [2] Bilash Kanti Bala, Fatimah Mohamed Arshad, Kusairi Mohd Noh, et al. “System dynamics”. In: *Modelling and Simulation* 274 (2017).
- [3] José Maria P Menezes Jr and Guilherme A Barreto. “Long-term time series prediction with the NARX network: An empirical evaluation”. In: *Neurocomputing* 71.16-18 (2008), pp. 3335–3343.
- [4] Olalekan Ogunmolu et al. “Nonlinear systems identification using deep dynamic neural networks”. In: *arXiv preprint arXiv:1610.01439* (2016).
- [5] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [6] Czako Zoltan and Sebestyen-Pal Gheorghe. “Self-Adaptive Artificial Intelligence Techniques Used for Anomaly Detection”. In: *PhD Thesis* (2023).
- [7] Thomas G Dietterich. “Ensemble methods in machine learning”. In: *International workshop on multiple classifier systems*. Springer. 2000, pp. 1–15.
- [8] Xuezheng Jiang et al. “An adaptive multi-class imbalanced classification framework based on ensemble methods and deep network”. In: *Neural Computing and Applications* 35.15 (2023), pp. 11141–11159.
- [9] Kishan G Mehrotra et al. “Ensemble Methods”. In: *Anomaly Detection Principles and Algorithms* (2017), pp. 135–152.
- [10] Charu C Aggarwal et al. *Outlier detection for temporal data*. Morgan & Claypool Publishers., 2014.
- [11] Andrew A Cook, Göksel Mısırlı, and Zhong Fan. “Anomaly detection for IoT time-series data: A survey”. In: *IEEE Internet of Things Journal* 7.7 (2019), pp. 6481–6494.
- [12] Pang Clement. *Anomaly Detection on Time Series Data*. <https://patents.justia.com/patent/20200034733>. U.S. Patent 20200034733. 2020.
- [13] Ane Blázquez-García et al. “A review on outlier/anomaly detection in time series data”. In: *ACM Computing Surveys (CSUR)* 54.3 (2021), pp. 1–33.
- [14] Paul-Adrian Călburean et al. “Prediction of 3-year all-cause and cardiovascular cause mortality in a prospective percutaneous coronary intervention registry: Machine learning model outperforms conventional clinical risk scores”. In: *Atherosclerosis* 350 (2022), pp. 33–40. ISSN: 0021-9150.
- [15] Thyago P Carvalho et al. “A systematic literature review of machine learning methods applied to predictive maintenance”. In: *Computers & Industrial Engineering* 137 (2019), p. 106024.

- [16] Sanda-Maria Avram and Mihai Oltean. “A Comparison of Several AI Techniques for Authorship Attribution on Romanian Texts”. In: *Mathematics* 10.23 (2022). ISSN: 2227-7390.
- [17] Ali Bou Nassif et al. “Machine learning for anomaly detection: A systematic review”. In: *Ieee Access* 9 (2021), pp. 78658–78700.
- [18] Warren S McCulloch and Walter Pitts. “A logical calculus of the ideas immanent in nervous activity”. In: *The bulletin of mathematical biophysics* 5.4 (1943), pp. 115–133.
- [19] Lennart Ljung et al. “Deep learning and system identification”. In: *IFAC-PapersOnLine* 53.2 (2020), pp. 1175–1181.
- [20] Furkan Elmaz and Özgün Yücel. “Data-driven identification and model predictive control of biomass gasification process for maximum energy production”. In: *Energy* 195 (2020), p. 117037. ISSN: 0360-5442.
- [21] Ronald J Williams and David Zipser. “A learning algorithm for continually running fully recurrent neural networks”. In: *Neural computation* 1.2 (1989), pp. 270–280.
- [22] Thomas M Breuel. “Benchmarking of LSTM networks”. In: *arXiv preprint arXiv:1508.02774* (2015).
- [23] Sima Siامي-Namini, Neda Tavakoli, and Akbar Siامي Namin. “The performance of LSTM and BiLSTM in forecasting time series”. In: *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 3285–3292.
- [24] Kishan G Mehrotra et al. *Anomaly detection*. Springer, 2017.
- [25] Seongkyu Yoon and John F MacGregor. “Fault diagnosis with multivariate statistical models part I: using steady state fault signatures”. In: *Journal of process control* 11.4 (2001), pp. 387–400.
- [26] James J Downs and Ernest F Vogel. “A plant-wide industrial process control problem”. In: *Computers & chemical engineering* 17.3 (1993), pp. 245–255.
- [27] Cory A. Rieth et al. “Additional Tennessee Eastman Process Simulation Data for Anomaly Detection Evaluation”. In: *Harvard Dataverse*. Harvard Dataverse, 2017.
- [28] Florian Schmidt. “Generalization in generation: A closer look at exposure bias”. In: *arXiv preprint arXiv:1910.00292* (2019).
- [29] Boris Hanin. “Which Neural Net Architectures Give Rise to Exploding and Vanishing Gradients?” In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio et al. Vol. 31. Curran Associates, Inc., 2018.
- [30] Daniel S Yeung et al. *Sensitivity analysis for neural networks*. Springer, 2010.
- [31] CALEB Braun et al. “Heavy-Duty Emissions Control Tampering in Canada”. In: *International Council Clean Transportation (ICCT) Report* (2022).
- [32] DIAS. *Diagnostic Anti-Tampering Systems*. <https://Dias-Project.Com>. Accessed: 2024-01-05.
- [33] Nima Amini and Qinqin Zhu. “Fault detection and diagnosis with a novel source-aware autoencoder and deep residual neural network”. In: *Neurocomputing* 488 (2022), pp. 618–633.
- [34] Zhang Wu and Qinan Wang. “A single CUSUM chart using a single observation to monitor a variable”. In: *International Journal of Production Research* 45.3 (2007), pp. 719–741.