

**CONTRIBUȚII LA SCREENINGUL DISLALIEI LA COPIII  
DE VÂRSTĂ ȘCOLARĂ MICĂ (6-10 ani)**

**Teză de doctorat – Rezumat**

pentru obținerea titlului științific de doctor la

Universitatea Politehnica Timișoara

în domeniul de doctorat Calculatoare și Tehnologia Informației

**autor philol.** Emilian-Erman Mahmut

conducător științific Prof.univ.dr.ing. Vasile STOICU-TIVADAR

iulie 2024

Teza de doctorat, intitulată „Contribuții la screeningul dislaliei la copiii de vârstă școlară mică (6-10 ani)”, constă în 114 pagini împărțite în 6 capitole, care includ 19 tabele, 65 de figuri și 77 de referințe bibliografice în sprijinul cercetării. Obiectivul principal al tezei de doctorat este dezvoltarea unui algoritm automatizat de screening pentru tulburările dislalice, axat pe următoarele aspecte-cheie:

**Independența lingvistică:** Algoritmul trebuie să funcționeze pe baza analizei valorilor numerice extrase din semnalul audio, astfel încât să fie aplicabil în diverse limbi. Acest lucru ar elimina nevoia de date de antrenament specifice limbii.

**Eficiență sporită pentru logopezi:** Pentru a sprijini activitatea laborioasă a logopezilor, soluția de screening se concentrează pe următoarele funcționalități esențiale:

**Segmentare riguroasă și consecventă:** Algoritmul trebuie să segmenteze în mod fiabil eşantioanele audio la nivelul fonemului țintă, asigurând o analiză consecventă a diferitelor pronunții.

**Clasificare precisă:** Algoritmul trebuie să clasifice cu acuratețe probele de vorbire pentru a identifica eventualele tulburări dislalice.

**Accesibilitate, eficacitate din punct de vedere al costurilor și al timpului:** Soluția ar trebui să valorifice resursele de tip open-source și să fie ușor accesibilă în zonele îndepărtate cu acces limitat la logopezi. Acest lucru asigură accesibilitatea și aplicabilitatea pe scară largă.

Obiectivul secundar este dezvoltarea unei infrastructuri de cercetare, care va servi două scopuri:

Colectarea și analiza datelor:

- Vor fi colectate și stocate date anonimizate referitoare la tulburările dislalice la subiecții de vârstă școlară timpurie.
- Aceste date pot fi utilizate pentru a genera rapoarte valoroase bazate pe diverse criterii,

inclusiv sexul, localizarea geografică și ratele de succes ale metodelor de terapie. Aceste informații pot fi extrase dintr-o bază de date relațională.

Platformă de cercetare viitoare: Infrastructura de cercetare poate servi drept fundație pentru viitoarele eforturi de cercetare axate pe tulburările de pronunție a sunetelor vorbirii. Cercetătorii pot profita de această platformă pentru a accesa date și pentru a colabora la dezvoltarea metodelor de evaluare și terapie a tulburărilor dislalice.

Per ansamblu, această cercetare urmărește să creeze un instrument de screening automatizat, independent de limbă, accesibil și eficient pentru a sprijini logopezii în detectarea potențialelor tulburări dislalice. Deși instrumentul de screening nu poate diagnostica definitiv dislalia, acesta poate servi ca o resursă valoroasă pentru logopezi în luarea deciziilor privind parcursul terapeutic și maximizarea impactului intervențiilor acestora prin:

- Îmbunătățirea eficienței detectării: screeningul automatizat poate accelera procesul de identificare a copiilor cu risc de dislalie.
- Facilitarea intervenției timpurii: detectarea timpurie permite intervenția terapeutică în timp util, ceea ce poate duce la îmbunătățirea rezultatelor pentru copiii cu tulburări dislalice.

Următoarea secțiune oferă o prezentare succintă a fiecărui capitol, subliniind contribuțiile cheie ale acestora la cercetarea generală.

**Capitolul 1: Pregătirea terenului** - Acest capitol stabilește bazele cercetării noastre. Capitolul prezintă provocările evaluării tulburărilor de pronunție a sunetelor vorbirii (SSD), explorează potențialul instrumentelor automatizate de screening și subliniază obiectivele specifice de cercetare care ne vor ghida investigația. Iată care sunt punctele esențiale:

- **Importanța terapiei logopedice și limitările metodelor actuale:** Terapia logopedică tradițională implică evaluarea cu metode neautomatizate, care pot fi consumatoare de timp și subiective. Există o nevoie de instrumente mai eficiente și mai obiective.
- **Accentul pe dislalie:** Cercetarea se concentrează pe dislalie deoarece aceasta afectează în primul rând producerea fizică a sunetelor vorbirii și poate fi analizată matematic.
- **Rolul cogniției și al achiziției limbajului:** Dezvoltarea vorbirii și a limbajului sunt interconectate. Dezvoltarea creierului și abilitățile cognitive influențează modul în care copiii dobândesc și utilizează limbajul.
- **Provocările segmentării sunetelor vorbirii:** Vorbirea este un flux continuu de sunete, iar identificarea fonemelor individuale (sunetele vorbirii) poate fi dificilă din cauza unor fenomene precum asimilarea fonologică (sunetele se influențează reciproc).
- **Obiectivele cercetării:**
  - o Dezvoltarea unui instrument de screening automatizat, independent de limbă pentru dislalie, utilizând analiza numerică a semnalelor audio.
  - o Acest instrument ar trebui să segmenteze eficient fonemele și să clasifice cu acuratețe eșantioanele de vorbire pentru a identifica eventualele dislalii.
  - o Soluția ar trebui să fie accesibilă, valorificând resursele de tip open-source pentru accesibilitate și o aplicabilitate mai largă.
  - o Ca obiectiv secundar, va fi construită o infrastructură de cercetare pentru colectarea datelor, analiza și cercetările viitoare privind tulburările de vorbire.

Capitolul oferă, de asemenea, o foaie de parcurs pentru disertație, care scoate în evidență conținutul fiecărui capitol și modul în care acesta contribuie la obiectivul general de cercetare.

**Capitolul 2: O fundație construită pe cunoștințe și analiză** - Un alt pas în descoperirea secretelor sunetelor vorbirii, acest capitol analizează literatura existentă privind evaluarea tulburărilor sunetelor vorbirii și metodele automatizate de screening. Capitolul identifică ideile de bază, limitările abordărilor actuale și zonele cu potențial de explorare suplimentară. Această analiză critică oferă o bază solidă pentru propriile noastre eforturi de cercetare.

Tulburările de pronunție a sunetelor vorbirii reprezintă o provocare obișnuită pentru copiii de vârstă preșcolară și școlară mică, afectând pronunția și putând duce la dificultăți sociale și academice. Intervenția timpurie este crucială, iar terapia logopedică asistată de calculator oferă instrumente promițătoare pentru logopezi.

În timp ce metodele tradiționale se bazează pe evaluări pe suport de hârtie, terapia logopedică asistată de calculator aduce avantaje precum feedback-ul imediat, activități care îmbracă forma unui joc sau a unei povești și colectarea datelor pentru monitorizarea progresului. Cu toate acestea, există preocupări cu privire la timpul petrecut în fața ecranului, accesibilitate și dependența excesivă de tehnologie, scoțând în evidență importanța automatizării în terapia logopedică ca supliment la terapia tradițională.

Screeningul automatizat al sunetelor vorbirii este un domeniu cheie pentru dezvoltarea instrumentelor automatizate pentru logopedie. Cercetătorii explorează modalitatea de cuantificare a pronunției acceptabile. Aceasta include definirea matematică a preciziei acustice necesare pentru o comunicare clară și determinarea intervalelor de variații naturale care nu împiedică înțelegerea cuvintelor. Teoria lui Shannon privind comunicarea ajută la reducerea decalajului dintre analiza obiectivă a sunetului și percepția umană subiectivă în timpul vorbirii.

Aplicațiile logopedice actuale se concentrează în principal pe exerciții pentru tulburările de pronunție. Limitările abordărilor existente sunt încă în curs de cercetare. Aceste limitări includ lipsa unei terminologii standardizate, natura complexă a evaluării tulburărilor de pronunție (necesitatea unor abordări interdisciplinare) și particularitățile pronunției specifice fiecărei limbi. În plus, cercetătorii explorează modul de integrare a capacității auditive și a auzului fonematic (distingerea sunetelor) în sistemele computerizate de terapie logopedică, analizând totodată efectele filtrării sunetului în lumea reală, în calea de la aparatul fonoarticulator al vorbitorului la cortexul auditiv al interlocutorului.

Cercetările recente au analizat entropia informațională ca modalitate de abordare a subiectivității inerente percepției pronunției. În plus, cercetătorii dezvoltă modalități ușor de utilizat de prezentare a rezultatelor screeningului, recurgând la reprezentări vizuale și tehnici de ludicizare (gamification) pentru a reduce presiunea psihologică resimțită de utilizator și a încuraja pronunția naturală în timpul evaluărilor.

Viitorul terapiei logopedice asistate de calculator oferă posibilități interesante. Cercetătorii explorează legăturile dintre calitatea comunicării și dislalie (tulburări ale sunetelor vorbirii), cu scopul de a valorifica datele de la screeningurile automatizate pentru a înțelege evoluția

tulburărilor de pronunție în rândul populației. Trebuie remarcat că aceste instrumente sunt concepute ca ajutoare pentru logopezi, oferind avantaje precum: screening eficient, reducerea erorilor și accesibilitate sporită în localitățile cu deficit de logopezi. Prin abordarea limitărilor actuale și explorarea unor noi căi, terapia logopedică asistată de calculator are potențialul de a deveni un instrument valoros în contracararea tulburărilor de pronunție.

Cercetarea noastră se aliază cu literatura existentă în ce privește următoarele aspecte ale extragerii caracteristicilor și algoritmilor:

- Amplitudinea drept caracteristică de discriminare a pronunției eronate: susținută de [1][2].
- Pronunția logopedului ca referință: pronunția logopedului servește drept referință pentru comparația cu segmentul de fonem analizat [3].
- Fonemul /r/ ca țintă inițială de validare: Fonemul românesc /r/ a fost ales pentru validarea inițială datorită caracteristicilor sale acustice (consoană lichidă laterală vibrantă) [1][2].
- Intervalul de vârstă vizat: aliniat cu grupul de vârstă vizat în [3].

Pentru a reduce decalajul subiectivitate-obiectivitate și a stimula motivația utilizatorilor, cercetarea noastră traduce rezultatele numerice ale screeningului în reprezentări izometrice în culori și nuanțe ale acestora.

Următorul capitol descrie un sistem de extragere a caracteristicilor și de clasificare a înregistrărilor audio, dezvoltat în conformitate cu subseturi ale obiectivelor de cercetare pentru a servi drept infrastructură de cercetare.

### **Capitolul 3: Crearea instrumentelor pentru extragerea și clasificarea caracteristicilor -**

Acest capitol reprezintă începutul etapei practice a cercetării și detaliază dezvoltarea unei aplicații software pentru extragerea și clasificarea caracteristicilor din înregistrările audio. Am dezvoltat o aplicație personalizată folosind C# în framework-ul .NET pentru a procesa datele audio și a valida obiectivele cercetării noastre. Explorarea diferitelor tehnici a deschis calea pentru selectarea celor mai robuste și eficiente metode pentru instrumentul nostru de screening automatizat.

Abordarea noastră inițială s-a axat pe succesiunea valorilor de amplitudine ca indicator al pronunției greșite. Am comparat pronunțiile subiecților cu cele ale unui logoped ca pronunții de referință. Am utilizat regresia polinomială pentru extragerea caracteristicilor și entropia informațională pentru clasificare, dar această metodă a prezentat limitări în ceea ce privește precizia și robustețea.

Pentru a îmbunătăți acuratețea, am explorat diferite ordine polinomiale pentru ajustarea liniilor de tendință, am suplimentat reprezentarea șirurilor de caracteristici și am încorporat informații despre amplitudine pentru a obține un set de caracteristici mai amplu. De asemenea, am ajustat rata de eșantionare pentru a controla nivelul de detaliu extras din semnalul audio.

Am rafinat și mai mult procesul de clasificare prin calculul coeficientului de determinare ( $R^2$ ) pentru a evalua gradul de potrivire al modelului cu segmentele de semnal vocal. În plus, am implementat înregistrarea automată a datelor pentru a accelera procesarea. Deși am

încorporat transformarea logaritmică și calculul distanței Levenshtein, aceste metode nu au adus îmbunătățiri semnificative.

Cercetarea a priorizat explorarea și menținerea flexibilității în tehnicile de extragere și clasificare a caracteristicilor. Această aplicație a servit drept bază pentru un sistem mai cuprinzător de screening al sunetelor vorbirii. Procesul iterativ de dezvoltare a permis îmbunătățirea continuă a tehnicilor de extragere și clasificare a caracteristicilor.

Pentru a asigura eficiența din punct de vedere al costurilor, securitatea datelor și scalabilitatea, am optat pentru o bază de date relațională. Stocarea criptată asigură protecția informațiilor personale. Structura modulară permite integrarea facilă a algoritmilor și a funcționalităților viitoare.

Arhitectura propusă oferă o platformă centralizată pentru gestionarea terapeuților logopedici (SLT), a subiecților, a probelor audio și a algoritmilor de analiză. Stocarea în cloud ar spori accesibilitatea și scalabilitatea datelor. Designul modular permite extinderea viitoare cu noi algoritmi și interfețe. Printre potențialele beneficii se numără îmbunătățirea gestionării datelor, accesibilitatea, ușurința în utilizare și eficiența din punct de vedere al costurilor/timpului.

Interfața cu utilizatorul (UI) permite implementarea unei serii de tipuri de utilizator (părinți, logopezi, cercetători) prin controlul accesului bazat pe roluri (RBAC). Fereastra principală acționează ca un hub central cu funcționalități adaptate fiecărui tip de utilizator. Designul user-friendly asigură ușurința de utilizare pentru utilizatorii cu diferite competențe tehnice.

Dezvoltarea aplicației .NET și a bazei sale de date a urmat o abordare practică, acordând prioritate funcționalității față de estetică pentru a îndeplini obiectivele imediate de cercetare. Deși nesofisticată aplicația rezultată a asigurat cu succes funcționalitățile esențiale necesare pentru validarea sau invalidarea ipotezelor de cercetare. În esență, capitolul 3 detaliază dezvoltarea unei infrastructuri de cercetare care servește ca element de bază pentru un sistem mai avansat de screening al sunetelor vorbirii.

Următorul capitol explorează identificarea unei abordări fiabile și omogene pentru segmentarea sunetelor vorbirii în unități relevante.

**Capitolul 4: Optimizarea metodelor de segmentare și clasificare** - O piatră de temelie importantă în identificarea unor fundamente solide pentru analiza sunetelor vorbirii, acest capitol se concentrează pe rafinarea algoritmilor de segmentare și clasificare. Capitolul explorează diverse abordări, analizează eficacitatea acestora și evidențiază limitările identificate. Acest proces iterativ a permis perfecționarea metodelor și un pas înainte către realizarea unui instrument de screening automatizat fiabil.

Acest capitol explorează un algoritm bazat pe entropie informațională pentru depistarea automatizată a pronunției consoanei /r/ la copiii vorbitori de limbă română (5-7 ani). Analiza a implicat probe audio de pronunție ale copiilor („RAFT”, „PARĂ”, „FAR”) înregistrate în timpul sesiunilor de terapie logopedică. Un logoped a evaluat corectitudinea și acuratețea fiecărei pronunții.

Algoritmul a analizat valorile pozitive ale amplitudinii semnalului audio pentru a genera o linie de tendință polinomială. Această linie de tendință a fost apoi caracterizată prin atribuirea literelor „A” (ascendent), „D” (descendent) sau „S” (stabil) pentru fiecare punct de maxim și de minim (crests and troughs). Matricele de tranziție și valorile entropiei informaționale au fost calculate pentru a evalua similaritatea dintre semnalele de referință (pronunția logopedului) și cele de test (pronunția subiectului). Un prag strict de similaritate a fost utilizat pentru diferența dintre valorile entropiei informaționale pentru a identifica eventualele erori de pronunție.

Analiza a atins rate ridicate de potrivire cu evaluările logopezilor pentru /r/ inițial (93,3%), cu o performanță optimă la ordinul polinomial 11 și un interval de 10% asignat literei S (stabil). Pentru /r/ medial și final am obținut o precizie mai scăzută, dar semnificativă (80.0% și, respectiv, 83.3%) cu configurații optime distincte. Trebuie remarcat că nu a existat o singură configurație a parametrilor care să producă în mod constant cele mai bune rezultate în toate situațiile (fonem în poziție inițială, mediană, finală).

Abordarea inițială a scos în evidență limitări. Deși s-a obținut o acuratețe ridicată în configurații specifice, au apărut neconcordanțe între diferitele setări ale parametrilor. Eforturile de a aborda aceste neconcordanțe prin parametri de aplicare suplimentari nu au dus la îmbunătățiri semnificative. Ordinele polinomiale mai mari de 11 și intervalele mai mari asignate literei S au produs rezultate mai slabe. Tehnicile de extragere a caracteristicilor pe baza funcției logaritmice sau pe baza valorilor de maxim sau minim relativ ale amplitudinii nu au îmbunătățit semnificativ rezultatele. Metodele de clasificare bazate pe valorile coeficientului de determinare ( $R^2$ ) și calculul distanței Levenshtein au fost, de asemenea, ineficiente. Capitolul 3 oferă o descriere detaliată a îmbunătățirilor aduse etapei de extragere a caracteristicilor. Din cauza acestor limitări, accentul cercetării s-a mutat pe dezvoltarea unui algoritm mai robust care utilizează învățarea automată pentru clasificare.

Următorul capitol descrie o nouă abordare bazată pe corelația încrucișată pentru etapa de segmentare, care se aliniază obiectivelor de cercetare stabilite. Corelarea încrucișată a abordat în mod eficient atât impactul asimilării fonologice, cât și problema duratei variabile a aceluiași enunț pronunțat de vorbitori diferiți. Această abordare a produs rezultate promițătoare și consecvente, care au fost optimizate în continuare prin implementarea clasificatorilor de învățare automată.

**Capitolul 5: Către o soluție promițătoare - Corelarea încrucișată și învățarea automată -** Acest capitol prezintă un progres semnificativ în cercetarea noastră. Capitolul explorează o nouă abordare în două etape care utilizează corelarea încrucișată pentru segmentare și învățarea automată (Machine Learning) pentru clasificare. Rezultatele demonstrează acuratețe și eficiență promițătoare, ceea ce ne apropie de realizarea unui instrument de screening automatizat robust și fiabil.

Etapa 1: Segmentare prin corelare încrucișată

Sistemul utilizează corelarea încrucișată pentru a segmenta fonemele individuale din înregistrările vocale. Această metodă gestionează în mod eficient variabilitatea lungimii pronunției și interacțiunile complexe dintre sunetele învecinate (asimilare fonologică).

Spre deosebire de metodele tradiționale, corelarea încrucișată nu necesită procesare manuală. Aceasta segmentează automat fonemul vizat din cuvinte întregi și generează caracteristici relevante pentru etapa următoare. Segmentarea fonemelor-țintă în poziție inițială, mediană și finală a fost realizată prin scrierea unor script-uri în Python. Segmentarea prin corelare încrucișată oferă mai multe avantaje:

- abordează complexitatea vorbirii naturale, inclusiv lungimile variabile ale pronunției și interacțiunile complexe ale fonemelor.
- asigură o segmentare consecventă și generează caracteristici valide pentru clasificare.
- automatizează procesul, eliminând necesitatea intervenției manuale.

Principalele constatări privind segmentarea prin corelare încrucișată (subcapitolele 5.1-5.3) sunt următoarele:

- Corelarea încrucișată a segmentat în mod eficient fonemul /r/ în diverse contexte (vocale, diftongi, consoane);
- Analiza a evidențiat influența asimilării fonologice asupra caracteristicilor fonemelor, subliniind importanța luării în considerare a sunetelor vecine (subcapitolul 5.1);
- Etapa de segmentare s-a dovedit robustă în gestionarea probelor audio din diferite cuvinte românești (subcapitolul 5.2).

Etapa 2: Clasificarea cu ajutorul învățării automate (Machine Learning)

Modelele de învățare automată au fost utilizate pentru a clasifica fonemele segmentate ca fiind corecte sau pronunțate greșit. Am implementat o selecție de algoritmi de învățare automată în Python, inclusiv Linear SVM [4][5], Decision Tree (CART) [6][7], AdaBoost [8][9], Gaussian Naive Bayes [10][11], Multi-Layer Perceptron (MLP) [12] și Logistic Regression [13][14]. Cercetarea s-a axat pe două modele: Linear SVM și Decision Tree (CART). Ambele modele au obținut o acuratețe ridicată (peste 97%) în clasificarea sunetului /r/ în diverse contexte românești (vocale, diftongi și consoane). În particular, clasificatorul Decision Tree a prezentat o acuratețe perfectă pentru toate metricile, în timp ce clasificatorul Linear SVM a prezentat o ușoară scădere a preciziei la un set de testare mai mare. Acest lucru sugerează că clasificatorul Linear SVM ar putea fi mai sensibil la cantitatea de date de antrenament pe care o primește.

Principalele constatări privind clasificarea prin învățare automată (subcapitolele 5.4-5.5) sunt următoarele:

- Atât modelul Linear Support Vector Machine (SVM), cât și modelul Classification and Regression Tree (CART) au obținut rezultate promițătoare, modelul CART obținând scoruri perfecte în anumite măsurători (subcapitolul 5.5).
- Precizia modelului Linear SVM s-a îmbunătățit ușor cu un set de date de formare mai mare, sugerând beneficiile potențiale ale creșterii numărului de date (subcapitolul 5.5).
- În timp ce modelul CART a arătat o performanță ridicată constantă, modelul Linear SVM a arătat sensibilitate la dimensiunea și distribuția datelor de antrenament, subliniind necesitatea optimizării suplimentare (subcapitolul 5.5).

Clasificarea prin învățare automată a contribuit la reducerea dificultății legate de compararea vocii femeii adulte (majoritatea logopezilor sunt femei adulte) cu vocile copiilor. Acest lucru

se datorează faptului că algoritmi de clasificare se concentrează în primul rând pe anvelopa spectrală și frecvențele formanțelor (timbrul) semnalului de vorbire, care sunt mai puțin influențate de anatomia coardelor vocale și de diferențele de înălțime a sunetului dintre vocile adulților și ale copiilor. Epocile de antrenare valorifică datele subiecților pentru a lua în considerare aceste variații și a îmbunătăți precizia generală a sistemului.

Cercetările ulterioare se vor concentra pe extinderea dimensiunii și diversității setului de date pentru a îmbunătăți performanța modelului pe date noi, nevăzute, permițând sistemului să clasifice o gamă mai largă de foneme pe lângă consoana /r/, integrând reguli specifice limbii pentru a oferi informații mai detaliate despre pronunțiile greșite, ajutând la intervenții specifice și validând eficacitatea sistemului în contexte clinice reale.

Din punctul de vedere al semnificației generale, această abordare în două etape demonstrează potențialul de dezvoltare a unui instrument automatizat de screening al tulburărilor de pronunție. Etapa de segmentare izolează eficient fonemele pentru o analiză ulterioară, în timp ce modelele de învățare automată oferă capacități de clasificare promițătoare.

**Capitolul 6: Concluzii generale** - Capitolul final al cercetării sintetizează întregul parcurs de cercetare. Capitolul rezumă principalele constatări, evidențiază contribuțiile cercetării la domeniul evaluării tulburărilor de pronunție și identifică piste de cercetare ulterioare.

Disertația descrie un sistem promițător de detectare automată a tulburărilor de pronunție la copii. Acest sistem în două etape își propune să sprijine logopezii și specialiștii în logopedie prin simplificarea procesului de screening.

Prima etapă abordează segmentarea fonemelor, procesul de izolare a sunetelor individuale din înregistrările audio. Am explorat diverse tehnici, dar am constatat că tehnica bazată pe corelarea încrucișată a excelat în segmentarea fonemelor, în special a consoanei /r/ în limba română. Această metodă gestionează în mod eficient variațiile în durata pronunției și interacțiunile complexe dintre sunetele învecinate. Spre deosebire de metodele tradiționale, corelarea încrucișată automatizează segmentarea, eliminând necesitatea procesării manuale.

A doua etapă abordează clasificarea. În etapa de clasificare modelele de învățare automată procesează fonemele segmentate pentru a identifica eventualele greșeli de pronunție. Disertația s-a axat pe două modele: clasificatorul Linear SVM și clasificatorul Decision Tree. Ambele modele au obținut o acuratețe impresionantă în clasificarea pronunțiilor corecte și incorecte ale consoanei /r/ în diverse cuvinte românești.

Această cercetare contribuie la domeniul evaluării tulburărilor de pronunție în mai multe moduri:

- Abordare nouă în două etape: Se propune o nouă abordare în două etape pentru depistarea automată a tulburărilor de pronunție, care valorifică funcția de corelare încrucișată pentru segmentare și învățarea automată pentru clasificare (capitolul 5). Această abordare oferă avantaje precum coerența, viteza și potențialul de independență lingvistică.
- Segmentare riguroasă și consecventă: O contribuție importantă a acestei cercetări este dezvoltarea unei etape de segmentare riguroase și consecvente care utilizează corelarea încrucișată (capitolul 5). Această abordare gestionează în mod eficient următoarele provocări:



- Asimilarea fonologică progresivă și regresivă: Corelarea încrucișată rezolvă problema influenței sunetelor învecinate asupra caracteristicilor fonemelor țintă, mai exact fenomenele cunoscute sub denumirea de asimilare progresivă și regresivă (capitolul 5). Acest lucru este fundamental pentru o segmentare precisă, deoarece fonemele adiacente își influențează reciproc pronunția.

- Durata variabilă: Metoda corelării încrucișate este robustă în gestionarea probelor audio (pronunții ale unor cuvinte românești) cu durate variabile (capitolul 5). Metoda elimină limitările metodelor tradiționale de segmentare, care prezintă dificultăți în cazul segmentării pronunțiilor cu durate diferite ale aceluiași cuvânt.

- Învățarea automată pentru provocările legate de timbrul vocii: Cercetarea demonstrează eficacitatea învățării automate în atenuarea provocărilor asociate cu variațiile de timbru între vocile adulților și ale copiilor (capitolul 5). Această constatare este valoroasă pentru dezvoltarea unor instrumente de screening care să evalueze riguros pronunțiile copiilor în pofida variațiilor naturale.

- Design open-source și modular: Accentul pus pe resursele de tip open-source și pe designul modular al aplicației software și a bazei de date promovează o accesibilitate sporită, posibilitatea de extindere în viitor și eficiența din punct de vedere al costurilor pentru logopezi și cercetători (capitolul 3).

În concluzie, cercetarea deschide posibilități de îmbunătățire a modului în care sunt detectate tulburările de pronunție la copii. Sistemul oferă mai multe avantaje:

- Eficiență pentru logopezi: procesul automatizat se traduce prin timpi de screening mai rapizi și un volum de muncă potențial redus pentru logopezi.

- Accesibilitate și eficiență din punct de vedere al costurilor: Utilizarea resurselor de tip open-source și un design modular promovează accesul mai larg și extinderea viitoare cu costuri suplimentare minime.

- Independență lingvistică: Accentul pus pe tehnicile de extragere a caracteristicilor independente de limbă deschide calea către un instrument cu potențial general.

Sistemul actual se concentrează în principal pe consoana /r/. Extinderea capacităților sistemului în vederea clasificării unei game mai largi de foneme va spori utilitatea sa generală. În plus, setul de date de antrenament utilizat pentru modelele de învățare automată trebuie să crească în mărime și diversitate pentru a ne asigura că instrumentul procesează corect date noi, nevăzute. Sunt necesare studii în lumea reală pentru a valida eficacitatea sistemului în identificarea tulburărilor de pronunție într-un cadru clinic.

Per ansamblu, această cercetare demonstrează potențialul unei abordări în două etape care utilizează învățarea automată pentru depistarea tulburărilor de pronunție.

Această teză contribuie în mod semnificativ la cunoștințele existente. Pe parcursul cercetării au fost publicate 11 lucrări științifice. Dintre acestea, șase lucrări sunt indexate în prestigioasa bază de date ISI/WOS. O altă lucrare a fost publicată într-o revistă indexată în BDI (Bibliographic Databases), iar alte patru lucrări indexate BDI sunt pe cale de a fi incluse în ISI/WOS. În plus, toate publicațiile menționează cercetătorul ca autor principal.

Iată articolele științifice care au rezultat din această cercetare:

1. E. E. Mahmut and V. Stoicu-Tivadar, "Current Challenges in the Computer-Based Assessment of Speech Sound Disorders," in 2018 IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania, May 17-19, 2018, pp. 431-435, doi: 10.1109/SACI.2018.8440970, WOS:000625278800072.
2. E. E. Mahmut and V. Stoicu-Tivadar, "A Speech Sound Disorder Screening System Database Structure," in Decision Support Systems and Education: Help and Support in Healthcare, J. Mantas, Z. Sonicki, M. Crisan Vida, K. Fister, M. Hagglund, A. Kolokathi, and M. Hercigonja Szekeres, Eds., ser. Studies in Health Technology and Informatics, vol. 255, Zagreb, Croatia, Oct. 15-16, 2018, pp. 185-189, European Federat Med Informat, 2018, doi: 10.3233/978-1-61499-921-8-185, WOS:000455957400036.
3. E. E. Mahmut, M. Della Ventura, D. Berian, and V. Stoicu-Tivadar, "Entropy-based Dyslalia Screening," in Health Informatics Vision: From Data via Information to Knowledge, J. Mantas, A. Hasman, P. Gallos, A. Kolokathi, M. S. Househ, and J. Liaskos, Eds., ser. Studies in Health Technology and Informatics, vol. 262, Athens, Greece, Jul. 5-7, 2019, pp. 252-255, doi: 10.3233/SHTI190066, WOS:000560388600065.
4. E. E. Mahmut, D. Berian, M. Della Ventura, and V. Stoicu-Tivadar, "Optimization of Entropy-Based Automated Dyslalia Screening Algorithm," in Digital Personalized Health and Medicine, L. B. Pape-Haugaard, C. Lovis, I. C. Madsen, P. Weber, P. H. Nielsen, and P. Scott, Eds., ser. Studies in Health Technology and Informatics, vol. 270, Geneva, Switzerland, Apr. 2020, pp. 357-361, European Federat Med Informat, 2020, doi: 10.3233/SHTI200182, WOS:000448144200075.
5. E. E. Mahmut, S. Nicola, and V. Stoicu-Tivadar, "Cross-Correlation Based Automated Segmentation of Audio Samples," in Importance of Health Informatics in Public Health During a Pandemic, J. Mantas, A. Hasman, M. S. Househ, P. Gallos, and E. Zoulias, Eds., ser. Studies in Health Technology and Informatics, vol. 272, Jul. 3-5, 2020, pp. 241-244, doi: 10.3233/SHTI200539, WOS:000630065600062.
6. E. E. Mahmut, S. Nicola, and V. Stoicu-Tivadar, "Cross-Correlation Based Automatic Segmentation of Medial Phonemes," in 2020 14th International Symposium on Electronics and Telecommunications (ISETC), Timisoara, Romania, Nov. 5-6, 2020, pp. 293-296, doi: 10.1109/isetc50328.2020.9301048, WOS:000612681000070.
7. E. E. MAHMUT, M. DELLA VENTURA, and V. STOICU-TIVADAR, "An Entropy-Based Computer Model for the Measurement of Phonetic Similarity: Dyslalia Screening in Early School-Age Children", Appl Med Inform, vol. 40, no. 1-2, pp. 15-23, Jun. 2018.
8. E. E. Mahmut, S. Nicola, and V. Stoicu-Tivadar, "Decision Tree Versus Linear Support Vector Machine Classifier in the Screening of Medial Speech Sounds: A Quest for a Sound Rationale," Studies in Health Technology and Informatics, vol. 309, pp. 73-77, Oct. 20, 2023, doi: 10.3233/SHTI230742.
9. E. E. Mahmut, S. Nicola, and V. Stoicu-Tivadar, "Support-Vector Machine-Based Classifier of Cross-Correlated Phoneme Segments for Speech Sound Disorder Screening," Studies in Health Technology and Informatics, vol. 294, pp. 455-459, May 25, 2022, doi: 10.3233/SHTI220500.
10. E. E. Mahmut, S. Nicola, and V. Stoicu-Tivadar, "Word-Final Phoneme Segmentation Using Cross-Correlation," Studies in Health Technology and Informatics, vol. 275, pp. 132-136, Nov. 23, 2020, doi: 10.3233/SHTI200709.
11. E. E. Mahmut, S. Nicola, and V. Stoicu-Tivadar, "A Computer-Based Speech Sound Disorder Screening System Architecture," Studies in Health Technology and Informatics, vol. 251, pp. 39-42, 2018, doi: 10.3233/978-1-61499-921-8-39.

## Bibliografie

- [1] Grigore, O., Grigore, C., Velican, V., Intelligent System for Impaired Speech Evaluation, Recent Advances in Circuits, Systems and Signals, Book Series: International Conference on Circuits Systems Signals, ISBN:978-960- 474-226-4, WOS:000290360900078, Malta, Sept. 2010, p. 365-368.
- [2] Grigore, O., Velican, V., Self-Organizing Maps For Identifying Impaired Speech, Politehnica University of Bucharest, 061071, Romania, Department of Applied Electronics and Telecommunications Faculty, in Advances in Electrical and Computer Engineering (AECE), Vol. 11, No. 3, 2011;
- [3] Pentiuc, Ș. G., Schipor, O. A., Gîză Belciug, F., Belciug, C. E., Nestor, M., Aplicație Logomon - Raport final Proiect TERAPERS, Sistem pentru terapia personalizată a tulburărilor de expresie lingvistică. Suceava: Universitatea Ștefan cel Mare, Facultatea de Inginerie Electrică și Știința Calculatoarelor, 2006;
- [4] Mahmut, E.E., Nicola, S., Stoicu-Tivadar, V., Support-Vector Machine-Based Classifier of Cross-Correlated Phoneme Segments for Speech Sound Disorder Screening, Stud Health Technol Inform., 294, p. 455-459, May 2022;
- [5] Kecman V., Learning and Soft Computing, Support Vector Machines, Neural Networks, and Fuzzy Logic Models, Cambridge, Massachusetts, The MIT PRESS, p. 148-166, 2001;
- [6] Kumar, P., Kumar, D., Decision tree classifier: a detailed survey, International Journal of Information and Decision Sciences, Vol. 12, No. 3, p. 246-269, Apr. 2020;
- [7] Mahmut, E.E., Nicola, S., Stoicu-Tivadar, V., Decision Tree versus Linear Support Vector Machine Classifier in the screening of medial speech sounds: a quest for a sound rationale, Stud Health Technol Inform., 309, p. 73-77, Oct. 2023;
- [8] Ding, Y., Zhu, H., Chen, R. and Li, R., An Efficient AdaBoost Algorithm with the Multiple Thresholds Classification, Appl. Sci., 12, 5872, Jun. 2022;
- [9] Wang, W. and Sun, D., The improved AdaBoost algorithms for imbalanced data classification, Information Sciences, Volume 563, p. 358-374, Jul. 2021;
- [10] Tieppo, E., Nievola, J.C. and Barddal, J.P., Adaptive learning on hierarchical data streams using window-weighted Gaussian probabilities, Applied Soft Computing, Vol. 152, Feb. 2024;
- [11] Chen, S., Webb, G.I., Liu, L. and Ma, X., A novel selective naïve Bayes algorithm, Knowledge-Based Systems, Vol. 192, March 2020;
- [12] Chatterjee, A., Saha, J. and Mukherjee, J., Clustering with multi-layered perceptron, Pattern Recognition Letters, Vol. 155, p. 92-99, Mar. 2022;
- [13] Hu, W., Qian, Y., Soong, F.K. and Wang, Y., Improved mispronunciation detection with deep neural network trained acoustic models and transfer learning based logistic regression classifiers, Speech Communication, Vol. 67, p. 154-166, Mar. 2015;
- [14] Siniscalchi, M.S., Combining speech attribute detection and penalized logistic regression for phoneme recognition, Neurocomputing, Vol. 93, p. 10-18, Sept. 2012;