



**UNIVERSITATEA „POLITEHNICA” din BUCUREȘTI**

**ȘCOALA DOCTORALĂ ETTI-B**

# **REZUMATUL TEZEI DE DOCTORAT**

**ARHITECTURĂ MULTIMODALĂ PENTRU  
PREDICȚIA PROCESELOR AFECTIVE UMANE**

Doctorand:  
**Ing. Mihai Gavrilescu**

Conducător Doctorat:  
**Prof. Dr. Ing. Nicolae Vizireanu**

**BUCUREȘTI 2019**

---

## **Cuprins**

|                                                                                 |           |
|---------------------------------------------------------------------------------|-----------|
| <b>CAPITOLUL 1 – INTRODUCERE .....</b>                                          | <b>1</b>  |
| <b>CAPITOLUL 2 - CONCEPTE PSIHOLOGICE UTILIZATE .....</b>                       | <b>1</b>  |
| <b>CAPITOLUL 3 - BAZE DE DATE UTILIZATE ȘI DETALII ARHITECTURALE GENERALE .</b> | <b>2</b>  |
| <b>CAPITOLUL 4 - ANALIZA CARACTERISTICILOR FACIALE.....</b>                     | <b>3</b>  |
| <b>CAPITOLUL 5 - ANALIZA CARACTERISTICILOR VOCALE .....</b>                     | <b>6</b>  |
| <b>CAPITOLUL 6 - ANALIZA CARACTERISTICILOR GRAFOLOGICE.....</b>                 | <b>9</b>  |
| <b>CAPITOLUL 7 - ANALIZA MULTIMODALĂ.....</b>                                   | <b>12</b> |
| <b>CONCLUZII .....</b>                                                          | <b>16</b> |
| <b>LISTA LUCRĂRILOR ORIGINALE.....</b>                                          | <b>18</b> |
| <b>BIBLIOGRAFIE.....</b>                                                        | <b>19</b> |

## CAPITOLUL 1 – INTRODUCERE

Datorită impactului major al emoțiilor în procesele cognitive esențiale și în interacțiunile umane, precum și datorită importanței personalității și temperamentului pentru înțelegerea oamenilor, ca un computer să poată interacționa eficient, natural și inteligent cu oamenii este necesar ca acesta să poată recunoaște și exprima astfel de procese afective. Aceste procese au un rol cheie în înțelegerea unor fenomene mai complexe, ca atenția, memoria și estetica, putând fi aplicate cu succes într-o serie de arii din educație [1], comunicare [2], divertisment [3], design [4], medicină [5] sau îmbunătățirea interacțiunilor om-computer [6]. Ca urmare a acestor necesități de cercetare a fost creată o nouă ramură a tehnologiei informației care poartă denumirea de computerizare afectivă (Affective Computing – AC), un domeniu interdisciplinar care se întinde de la tehnologia informației și știința computerelor, la psihologie și științe cognitive. Este evident că, dat fiind progresul tehnologic, acest domeniu va fi unul din ce în ce mai cercetat. AC începe să devină un domeniu activ al tehnologiei informației, explorând moduri prin care se pot dezvolta dispozitive pentru predicția emoțiilor sau a altor aspecte psihologice ale oamenilor precum și a modului în care aceste dispozitive trebuie să reacționeze în funcție de aspectele psihologice determinate, dar și un anume „affective mediation“, un ansamblu de metode propuse pentru a facilita comunicarea între indivizi care își dezvăluie emoțiile [7]. Aceste direcții de cercetare s-au dezvoltat cu precădere în ultimul deceniu și se alătură altor cercetări în domeniul tehnologiilor asistive pentru îmbunătățirea zonei, neglijată până acum, a comunicării om-computer și a celei mediate de computer bazate pe procese afective [7]. În această teză propunem arhitecturi neinvazive care să permită predicția emoțiilor, personalității, temperamentului și stărilor emoționale. Teza abordează, în primă fază, o metodologie unimodală prin analiza separată a trei tipuri de caracteristici (faciale, vocale, grafologice) culminând cu analiza multimodală în care sunt determinate modurile de fuziune optimă pentru a asigura o acuratețe ridicată.

## CAPITOLUL 2 - CONCEPTE PSIHOLOGICE UTILIZATE

Există o gamă largă de moduri prin care procesele afective pot fi definite și evaluate, în această teză concentrându-ne pe cele mai eficiente și larg utilizate practic. Pentru **definirea emoțiilor**, plecăm de la studiul lui Clynes [8] care întărește ideea purității unor stări „sentice” primare și ne bazăm pe studiul lui Paul Ekman [9] care determină șase emoții primare, utilizate și în acest studiu: *furie*, *surprindere*, *frică*, *fericire*, *tristețe* și *dezgust*. Folosim Sistemul de Codare a Activității Faciale (Facial Action Coding System – FACS) [9] pentru analiza microexpresiilor sub forma unor unități de activare (Action Unit – AU) cu principalul avantaj de a determina emoțiile reale, chiar dacă subiectul încearcă să simuleze alte emoții [9]. În ultima revizie FACS [10], fața este împărțită în 46 AU-uri și 6 niveluri de intensitate pentru fiecare AU [10]. Pentru **definirea personalității**, utilizăm Modelul celor 5 factori de personalitate (Five-Factor Model – FFM) [11] care propune 5 dimensiuni de personalitate fundamentale după cum urmează [11]: *Deschiderea spre experiență (O)*, *Conștiinciozitate (C)*, *Extroversie (E)*, *Agreabilitate (A)*, *Neuroticism (N)*. FFM este utilizat cu succes într-o varietate de domenii, de la dezvoltare personală și profesională, până la predicția tulburărilor de personalitate, a persoanelor sub influența substanțelor [12] sau a bolilor fizice cu simptome asociate cu evoluția dimensiunilor FFM [13]. Pentru **definirea temperamentului**, utilizăm modelul Orientării Fundamentale a Relațiilor Interpersonale (Fundamental Interpersonal Relations Orientation – Behavior - FIRO-B) [14] bazat pe trei nevoi interpersonale principale care determină un individ să acționeze sau nu atunci când este parte a unui grup:

*Deschidere, Control și Includere* [14]. Aceste trei nevoi principale sunt evaluate prin intermediul unui chestionar specific, de la 0 la 9, pe două coordonate: *Expressed* (Exprimare) și *Wanted* (Dorință) [14]. Rezultatele sunt cartografiate în 5 temperamente fundamentale: melancolic, coleric, sangvinic, supin și flegmatic. FIRO-B este utilizat cu precădere pentru construirea de echipe performante, consiliere de carieră și dezvoltare profesională [15]. Pentru **definirea stărilor emoționale** utilizăm Scara de Analiză a Depresiei, Anxietății și Stresului (Depression Anxiety Stress Scale – DASS) [16]. DASS propune un chestionar de auto-analiză (Self-Analysis Questionnaire - SAQ) care permite evaluarea a 14 elemente asociate depresiei, anxietății și stresului [16]. Rezultatul obținut reprezintă nivelul de severitate a stării emoționale, de la *Normal* la *Sever Extrem*. DASS este utilizat cu succes pentru determinarea severității celor 3 stări emoționale [16], dar și pentru evaluarea psihologică a angajaților [16], învățare adaptivă [17] sau diagnoza bolilor fizice cu una sau mai multe astfel de stări emoționale acute drept simptome [18].

### CAPITOLUL 3 - BAZE DE DATE UTILIZATE ȘI DETALII ARHITECTURALE GENERALE

Ne dorim să comparăm arhitecturile propuse în această teză cu alte metode din literatură, astfel că utilizăm o serie de baze de date menționate în tabelul de mai jos.

| Proces afectiv          | Analiză facială                                                                | Analiză vocală                                                                                                | Analiză grafologică | Analiză multimodală                                                |
|-------------------------|--------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|---------------------|--------------------------------------------------------------------|
| <b>Emoție</b>           | CK [19], CK+ [20], JAFFE [21], MMI [22], RU-FACS [23], Belfast [24], AFEW [25] | AVIC [28], eINTERFACE [29], McGilloway [30], Belfast structurată [24], Belfast naturalistică [24], SALAS [30] | -                   | eINTERFACE [29], AVIC [29], Belfast naturalistică [24], SALAS [30] |
| <b>Personalitate</b>    | -                                                                              | SSPNet SPC [31]                                                                                               | -                   | -                                                                  |
| <b>Temperament</b>      | -                                                                              | SSPNet SPC [31]                                                                                               | -                   | -                                                                  |
| <b>Stare emoțională</b> | ANUStressDB [26] (stres), AVEC2014 [27] (depresie)                             | SUSAS [32] (stres), AVEC2014 [27] (depresie), France et al. [34] (depresie)                                   | EMOTHAW [33]        | AVEC 2014 [27] (depresie)                                          |

Întrucât este necesară și o bază de date care corelează toate cele trei tipuri de caracteristici analizate cu cele patru procese afective evaluate și nu există baze de date atât de complete în literatură, am creat propria bază de date numită *MENTAL* (**M**ultimodal - **E**motion perso**N**ality Temperament - **A**ffective state **L**abeling), implicând 128 subiecți caucazieni (64 bărbați, 64 femei), între 18 și 35 de ani, aleși pentru păstra o distribuție uniformă a celor 5 temperamente și pentru a expune toate nivelurile DASS. Am colectat exemplare scrise de mână și înregistrări audio-video ale feței din poziție frontală în situații în care nicio emoție nu a fost indusă, precum și în situații în care fiecare dintre cele șase emoții primare a fost indusă. În fiecare astfel de sesiune colectăm și rezultatele obținute în urma completării chestionarelor: DEQ (Discrete Emotions Questionnaire), FFM, FIRO-B și SAQ. Rezultatele sunt evaluate de specialiști pentru a exclude exemplarele care nu sunt corect completate. Pentru inducerea celor 6 emoții utilizăm videoclipuri din baza de date LIRIS-ACCEDE [35]. *MENTAL* este împărțită într-un *set de date controlat* (*Dataset – controlled - DSC*) unde eşantioanele sunt colectate când o emoție este indusă sau textul este impus și *set de date aleator* (*Dataset – random - DSR*) unde eşantioanele sunt colectate când nicio emoție nu este indusă sau textul este liber ales de subiect.

Deși există o gamă largă de tipuri de rețele neuronale care ar putea fi utilizate, în această teză evaluăm exclusiv rețele neuronale de tip feed-forward (Feed-Forward Neural Network – FFNN). Pentru majoritatea sistemelor bazate pe FFNN nu este nevoie de mai mult de 2 niveluri ascunse [36], astfel că testăm arhitecturile folosind doar FFNN-uri cu un singur nivel ascuns sau cu 2 niveluri ascunse. Ca funcții de activare, evaluăm *tanh*, *sigmoid*, *ReLU* și *softmax*. Ca metodă de antrenare folosim algoritmul de propagare înapoi (*backpropagation*) [37] și utilizăm metoda gradientului descendent pentru a optimiza ponderile și bias-urile și a minimiza eroarea

relativă medie la o valoare mai mică de 0,02. Folosim metoda de inițializare a ponderilor Nguyen-Widrow [38] pentru a le împărți egal în starea inițială. Pentru a crea modele unimodale consistente care să permită o fuziune multimodală ușoară, toate arhitecturile propuse au o structură cu 3 niveluri unică în literatură: *un nivel inferior* în care caracteristicile sunt evaluate și clasificate, *un nivel intermediar* în care se construiește, pe baza clasificărilor din nivelul inferior, o matrice a caracteristicilor și *un nivel superior* în care o structură bazată pe FFNN-uri analizează matricea din nivelul intermediar și realizează predicția procesului afectiv analizat. Întrucât conțin FFNN-uri, toate arhitecturile propuse funcționează în două faze: faza de antrenare și faza de testare. În **faza de antrenare** un set de eșantioane sunt furnizate nivelului inferior care le normalizează și aplică o serie de transformări în funcție de tipul analizei (facială, vocală sau grafologică), realizând clasificarea caracteristicilor analizate. Rezultatele clasificării sunt trimise nivelului intermediar care construiește o matrice a caracteristicilor. În nivelul superior, o structură bazată pe FFNN-uri este antrenată prin algoritmul de propagare înapoi pe baza matricei caracteristicilor cu scopul de a oferi aceleași rezultate ca cele obținute prin completarea chestionarelor. Antrenarea ia sfârșit când eroarea este minimizată (sub 0,02) sau când eșantioanele de antrenare sunt epuizate. **Faza de testare** urmează pași similari. Eșantioanele sunt normalizate și le sunt aplicate transformări specifice, urmând clasificarea caracteristicilor analizate, rezultatul clasificărilor fiind trimis către nivelul intermediar unde se construiește matricea caracteristicilor. În nivelul superior, o structură anterior antrenată bazată pe FFNN-uri analizează matricea construită și oferă un rezultat în intervalul [0;1] reprezentând probabilitatea de apariție a procesului afectiv analizat. Utilizăm un clasificator bazat pe reguli (Rule-based classifier – RBC) pentru a furniza rezultatul final, astfel: dacă același rezultat este furnizat de fiecare FFNN timp de 5s consecutive, acesta este rezultatul final; altfel, RBC-ul marchează rezultatul final ca *Nedefinit*.

#### CAPITOLUL 4 - ANALIZA CARACTERISTICILOR FACIALE

Pentru analiza caracteristicilor faciale folosim abordarea dinamică și modelul FACS. Din cele 46 AU-uri prezente în ultima revizie FACS analizăm AU1, AU2, AU4-AU7, AU43 și AU45 din partea superioară a feței și AU8-AU20, AU22- AU28 din partea inferioară a feței. Pentru a îmbogăți informația emoțională din zona obrajilor analizăm și AU33, AU34 și AU35. Arhitectura este prezentată în Fig. 4.1. În primă etapă, stabilim un model statistic pentru culoarea pielii folosind o metodă similară cu cea prezentată în [39], clasificând pixelii în pixeli faciali și pixeli non-faciali pentru detecția feței [39]. Aplicăm filtre pentru eliminarea zgomotului [40] și pentru operații morfologice (eroziune, dilatare și umplerea golurilor) [41], obținând astfel fețe *candidat*. După detecția fețelor *candidat* determinăm poziția ochilor și a gurii [42], realizând și o scalare a feței prin calculul distanței interoculare. Extragem apoi caracteristicile faciale prin intermediul Modelelor de formă active (Active Shape Models – AAM) [43] și folosind Analiza Componentei Principale (Principal Component Analysis – PCA), producem un model parametrizat care descrie fețele utilizate în faza de antrenare și estimează fețe noi [44]. Învățarea corelațiilor dintre parametrii modelului și fața *candidat* se realizează prin metoda descendentului extrem (Steepest Descent - SD) folosind matrici Jacobiene [43][45]. Pentru identificarea caracteristicilor faciale creăm modele de benzi cu dimensiune egală centrate în jurul unor puncte de reper determinate pentru fiecare caracteristică facială [46]. Folosim gradientul stohastic descendent (Stochastic Gradient Descent – SGD) pe care îl modificăm în direcția opusă până atingem minimul, obținând astfel un set de parametri non-rigizi. Utilizăm clasificatori bazați pe mașini cu vector suport (Support Vector Machine – SVM) cu 6 clase [47] care primesc

drept intrare acești parametri non-rigizi și sunt antrenați să determine intensitatea fiecăreia dintre cele 31 AU-uri analizate. Realizând teste prin combinarea bazelor de date folosind CK+ [20], MMI [21], JAFFE [22], Belfast [24], AFEW [25] și RU-FACS [23], metoda propusă clasifică toate AU-urile cu peste 90% acuratețe. Pentru fiecare cadru din eșantionul video, fiecare AU este clasificat de la nivelul O la E, rezultatul fiind normalizat în intervalul [0;1]. Pe baza acestor clasificări, în nivelul intermediar construim o Matrice Facială (Face Matrix - FM), fiecare linie  $n$  conținând scorurile de clasificare pentru AU-uri evaluate în cadrul  $n$ .

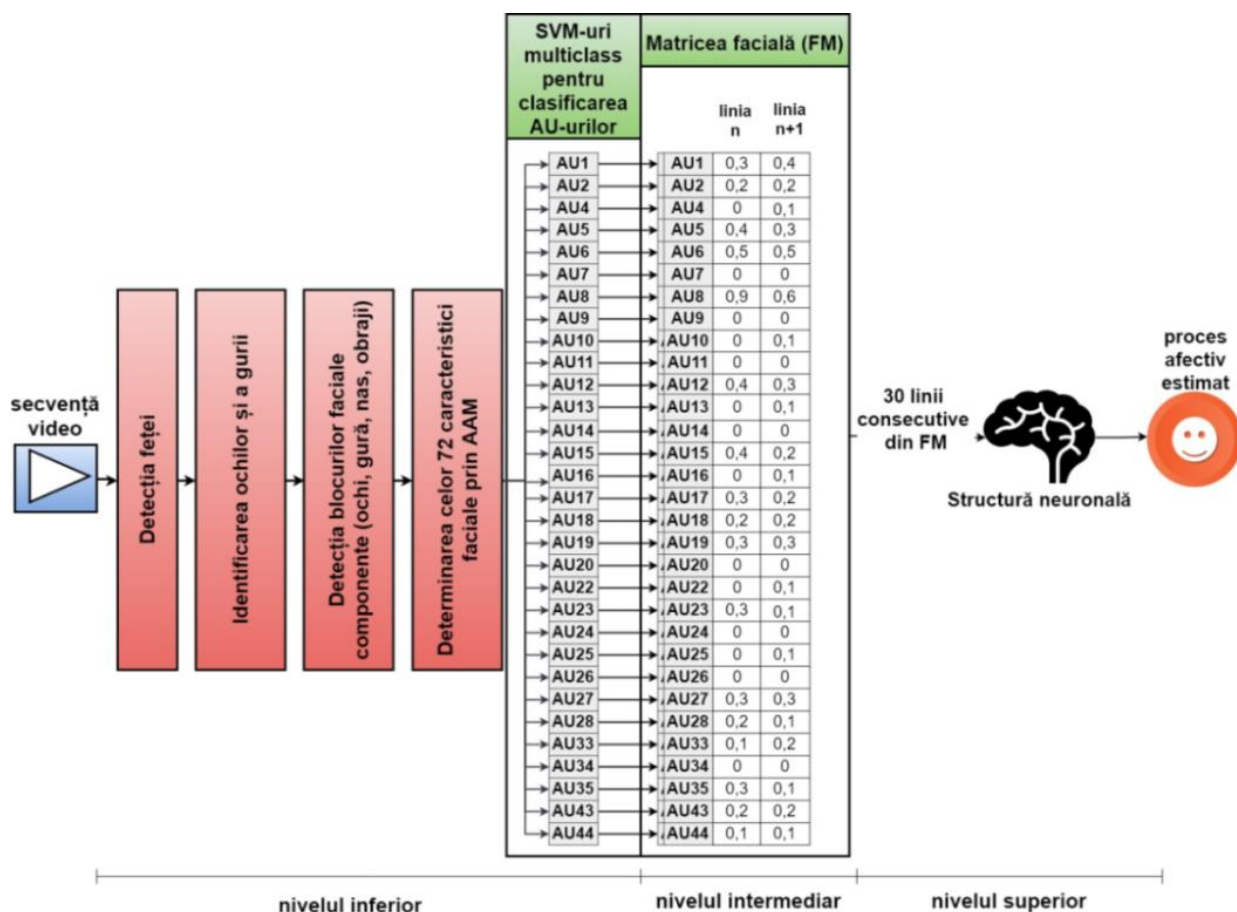


Fig. 4.1 Arhitectura generală pentru predicția proceselor afective prin analiza caracteristicilor faciale

Pentru **predicția emoțiilor pe baza analizei faciale** structura neuronală determinată optimă este compusă din 6 FFNN-uri dedicate fiecărei emoții, arătând astfel că emoțiile fiind astfel slab corelate din punct de vedere facial. Cea mai mare acuratețe este obținută când sistemul este antrenat și testat folosind DSC (87,91%), în timp ce, când folosim DSC doar pentru antrenare și testăm folosind DSR, acuratețea scade cu maxim 2%, arătând că DSC adaugă mai multă valoare în faza de antrenare ceea ce face sistemul practic întrucât expresiile faciale pot fi evaluate în condiții aleatoare fără a afecta considerabil acuratețea sistemului. În acest caz, cel mai relevant practic, timpul maxim necesar pentru a converge la un rezultat este 33s, semnificativ mai rapid decât timpul necesar completării DEQ (12,1 minute), metoda putând fi folosită pentru înlocuirea acestui chestionar. Cea mai mare acuratețe este obținută pentru „fericire”, urmată de „dezgust”, „tristețe” și „surprindere”, iar pentru „furie” și „frică” acuratețea este mai mică de 86%. Analizăm AU-urile prezente la intensitate mare atunci când o emoție este prezisă corect și determinăm o serie de conexiuni între AU-uri și emoții, convergente cu cele prezente în FACS [20], arătând că arhitectura este robustă. Folosind aceste corelații pentru predicția emoțiilor respective direct, fără a mai fi procesate de structura neuronală atunci când AU-urile corespondente sunt clasificate cu intensitate mare, acuratețea sistemului crește cu până la 5%. Comparând metoda propusă cu alte metode din literatură, aceasta oferă cea mai mare acuratețe. Această cercetare a fost publicată în [48] și extinsă în [49] pentru

predicția oboselii și în [50] și [51] prin fuziunea sa cu analiza poziției corpului și gesturilor pentru predicția emoției, conducând la o creștere a acurateței de până la 7% comparativ cu analiza exclusivă a expresiilor faciale.

Pentru **predicția dimensiunilor FFM pe baza analizei faciale**, structura neuronală optimă este reprezentată de un singur FFNN care modelează toate cele 5 dimensiuni. Din punct de vedere facial, cele 5 dimensiuni FFM sunt, astfel, puternic corelate. Obținem acuratețea cea mai mare atunci când DSC este utilizat atât pentru antrenare cât și pentru testare (77,4%), în timp ce, dacă păstrăm DSC pentru antrenarea sistemului și îl testăm folosind DSR, acuratețea scade cu mai puțin de 0,5%. Timpul maxim necesar pentru predicția dimensiunilor FFM este de 150s, considerabil mai mic decât cel necesar completării chestionarului FFM (21,2 minute), argumentând că arhitectura propusă îl poate înlocui, înlăturând subiectivitatea asociată cu acesta și permițând monitorizarea în timp real. Cea mai mare acuratețe o obținem pentru „Extroversie” (83,5%) și „Deschiderea spre Experiență” (79,94%), celelalte dimensiuni fiind prezise cu o acuratețe mai mică de 77%. Determinăm și aici un set de corelații între dimensiunile FFM și AU-urile analizate, fiind primul studiu din literatură care realizează o astfel de analiză. Integrarea acestor corelații în arhitectura propusă conduce la o creștere a acurateței cu până la 9%. Comparând alte metode din literatură cu arhitectura propusă, obținem acuratețe cu peste 5% mai mare decât metodele bazate pe Exaggeration Mapping (EM) [52] și rețele neuronale artificiale (Artificial Neural Networks - ANN) [53], atingând performanțe similare celor obținute de metoda bazată pe CERT [54]. Această cercetare a fost publicată în [55] și extinsă în [56] pentru predicția celor 16 factori de personalitate.

În cazul **predicției temperamentului pe baza analizei faciale**, acuratețea cea mai mare este obținută când folosim FFNN-uri dedicate pentru fiecare dintre cele 5 temperamente. Astfel, temperamentele nu sunt corelate facial, analiza lor combinată ducând la scăderea performanței sistemului. Obținem cea mai mare acuratețe (80,78%) când folosim DSC atât pentru antrenare cât și pentru testare, acuratețea scăzând cu peste 13% atunci când DSR este folosit în ambele faze. Și în acest caz, prin păstrarea DSC doar pentru antrenare, acuratețea scade cu mai puțin de 1%. Timpul maxim necesar predicției temperamentului este de 95s, considerabil mai rapid decât cel necesar completării chestionarului FIRO-B (14,1 minute). Temperamentul sangvinic este prezis cu cea mai mare acuratețe (85,12%), urmat de supin (83%) și melancolic (80,9%). Identificăm, pentru prima dată în literatură, un set de corelații între tipurile de temperament și FACS. Modificând arhitectura propusă pentru a cuprinde aceste corelații, obținem o creștere cu 7% a acurateței. Comparând arhitectura propusă cu metoda bazată pe ANN [53], observăm că arhitectura propusă oferă o acuratețe cu până la 5% mai mare.

Pentru **predicția stărilor emoționale pe baza analizei faciale**, structura optimă este aceea în care folosim un singur FFNN care modelează toate cele trei stări emoționale arătând că cele trei stări emoționale sunt puternic corelate din punct de vedere facial. Cea mai mare acuratețe este obținută prin utilizarea DSC atât pentru antrenare cât și pentru testare: 81,47% (stres), 79,83% (depresie), 68,65% (anxietate). Similar arhitecturilor anterioare, dacă menținem DSC pentru antrenarea sistemului și folosim DSR pentru testare, acuratețea scade cu mai puțin de 2%, ceea ce arată că inducerea emoțiilor este necesară doar în faza de antrenare, pentru testarea sistemului înregistrări colectate în condiții în care emoții nu au fost stimulate fiind la fel de eficiente. Cea mai mare acuratețe pentru toate cele trei stări emoționale o obținem pentru nivelurile *Normal* și *Sever Extrem*, în timp ce pentru celelalte niveluri intermediare (*Slab*, *Moderat* și *Sever*) acuratețea este mai mică, fiind cel mai des confundate cu nivelurile vecine. Arhitectura oferă, astfel, posibilitatea de a determina dacă un subiect este afectat

de depresie, anxietate și stres sau nu, însă pentru o diferențiere mai bună a severității stărilor emoționale alte metode trebuie explorate, cum ar fi includerea altor AU-uri sau analiza multimodală. Timpul maxim necesar pentru predicția nivelurilor DASS este de 64s, semnificativ mai rapid decât timpul necesar completării SAQ, sistemul putând înlocui chestionarul specific dacă acuratețea predicției nivelurilor intermediare este îmbunătățită. Realizăm primul studiu din literatură care identifică corelații între AU-urile analizate și cele trei stări emoționale. Modificând arhitectura pentru a cuprinde aceste corelații, acuratețea crește cu 5%. Pentru evaluarea stresului, arhitectura propusă oferă o acuratețe mai mare comparativ cu alte metode din literatură atunci când sunt testate pe bazele de date ANUStressDB [26] și MENTAL. Pentru evaluarea depresiei, arhitectura propusă oferă o acuratețe cu 1,5% mai mare decât metoda care utilizează modele binare locale în trei planuri ortogonale (Local Binary Patterns in Three Orthogonal Planes - LBP-TOP) și SVM [57]. Această cercetare a fost publicată în [58].

## CAPITOLUL 5 - ANALIZA CARACTERISTICILOR VOCALE

Deși larg utilizat în diferite tipuri de cercetări în ultimele decenii, semnalul vocal este încă un semnal complex din cauza variației sale lingvistice și semiotice, precum și celei legate de emoție și alte procese afective umane [59], dar și gradului de afectare cu zgomot și reverberație, majoritatea arhitecturilor fiind testate folosind rostiri neutre, înregistrate în studio [59], în absența acestor condiții rezultatele fiind mult mai puțin performante. La acestea se adaugă și faptul că astfel de sisteme presupun alegerea unui set de caracteristici optime dintr-un spectru larg, selectarea eficientă a caracteristicilor fiind complexă și deseori realizată prin TAE. Din cauza acestor motive, analiza semnalului vocal pentru predicția proceselor afective este una dificilă.

În această teză, datorită faptului că sunt larg folosite în cercetări similare [60], precum și datorită robusteții și abilității de a fi clasificate cu acuratețe ridicată folosind metode care nu sunt costisitoare din punct de vedere computațional [60], utilizăm următoarele caracteristici vocale: conturul pitch-ului (numărul de vârfuri pe secundă – *PMAXCOUNT*, media valorilor maxime – *PMAXAVG*, varianța valorilor maxime – *PVAR*, gradientul – *PGRAD*), conturul intensității (numărul de vârfuri pe secundă – *IMAXCOUNT*, media valorilor maxime – *IMAXAVG*, varianța valorilor maxime – *IVAR*, gradientul – *IGRAD*), rata de vorbire (Speech rate – *SR*), rata de pauze (Pause rate – *PR*, Rata trecerilor prin zero (Zero-crossing rate – *ZCR*), energia de durată scurtă (poziția, valoarea medie – *EAVG*, deviația standard – *ESTDEV*), jitter-ul (*JIT*), shimmer-ul (*SHIM*), frecvența fundamentală (*FFM*) și 33 de coeficienți MFCC (notați *MFCC0* – *MFCC32*). Fig. 5.1 prezintă structura nivelurilor inferior și intermediar. Toate arhitecturile propuse în acest capitol sunt bazate pe aceste 2 niveluri, diferența constând în structura neuronală a nivelului superior. **Nivelul inferior** are scopul de a prelua semnalul vocal reprezentând rostirea subiectului analizat și a clasifica un set de caracteristici vocale. Acesta este împărțit în 3 blocuri componente. În **blocul de normalizare** utilizăm scăderea spectrală pentru reducerea zgomotului [61], apoi normalizăm energia folosind normalizare cepstrală medie (Cepstral Mean Normalization - CMN) cu rețele bayesiene (Bayesian Networks - BN) [62] și, în final, normalizăm pitch-ul prin intermediul metodei bazate pe semitonuri propusă în [63]. Utilizăm și metodele pe care le-am publicat în [64], [65] și [66] pentru îmbunătățirea semnalului vocal. În **blocul pentru analiza și extracția caracteristicilor vocale la nivelul rostirii** analizăm ultimele 7s din rostire de fiecare dată când un cadru nou este furnizat drept intrare, caracteristicile vocale determinate în acest bloc fiind: *PMAXCOUNT*, *PVAR*, *PGRAD*, *IMAXCOUNT*, *IVAR*, *SR*, *PR*, *STE*, *EAVG*, *ESTDEV*, *JIT*, *SHIM* și *FF*. În **blocul pentru analiza și extracția caracteristicilor vocale la nivelul**



**cadrlui**, determinăm următoarele caracteristici pentru fiecare cadru nou furnizat drept intrare: *PMAXAVG*, *IMAXAVG*, *ZCR*, *JIT*, *SHIM*, *MFCC0*, ..., *MFCC32*. Pentru fiecare cadru din înregistrarea vocală clasificăm, astfel, 50 caracteristici vocale care sunt trimise blocului intermediar și stocate în Matricea Vocală (Speech Matrix - SM) care va conține valorile normalizate pentru fiecare caracteristică în parte și pentru fiecare cadru din rostire. La fiecare 30 de cadre consecutive neprocesate din SM, acestea sunt furnizate nivelului superior pentru a realiza predicția procesului afectiv analizat.

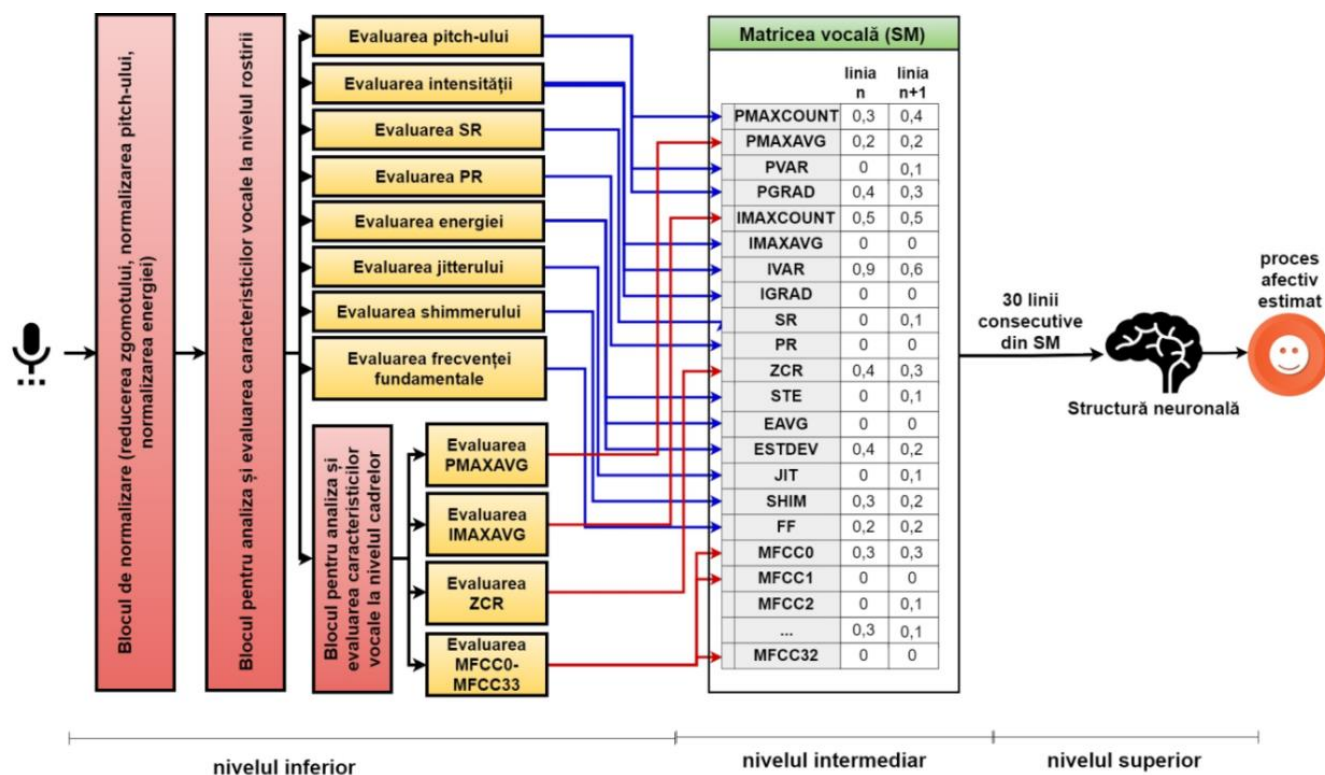


Fig. 5.1 Arhitectura generală pentru predicția proceselor afective prin analiza caracteristicilor vocale

Pentru **predicția emoțiilor pe baza analizei vocale**, structura optimă o reprezintă un singur FFNN utilizat pentru a modela toate cele 6 emoții. Comparativ cu analiza facială unde emoțiile sunt discrete, din punct de vedere vocal emoțiile sunt puternic corelate. Obținem o acuratețe de 80,75% atunci când sistemul este antrenat și testat folosind DSC, rezultate cu până la 12% mai mici fiind obținute când folosim exclusiv DSR pentru antrenare și testare. Dacă păstrăm DSC pentru antrenare și testăm folosind DSR, acuratețea scade cu mai puțin de 1,5%. Timpul maxim necesar predicției emoției în acest caz considerat cel mai practic este de 78s, semnificativ mai rapid decât DEQ (12 minute), ceea ce face ca arhitectura propusă să fie atractivă pentru înlocuirea chestionarului, dat fiind avantajul de a putea fi utilizată în timp real. Obținem cea mai mare acuratețe pentru „fericire” (88,1%), „tristețe” (81,8%), „frică” (80,5%) și „furie” (77,6%). Determinăm, pentru prima dată în literatură, o serie de corelații între emoții și caracteristicile vocale analizate care, utilizate pentru a îmbunătăți sistemul, conduc la o creștere a acurateței cu până la 6%. Comparând arhitectura propusă cu alte metode din literatură, aceasta oferă o acuratețe mai mare decât metodele bazate pe optimizarea ansamblurilor de particule (Particle Swarm Optimization – PSO) [67], modele bazate pe combinări Gaussiene (Gaussian Mixture Models – GMM) [68][69], ANN și Modele Markov Ascunse (Hidden Markov Models – HMM) [70], SVM și clasificatori de tip vecinul cel mai apropiat (K-Nearest Neighbors - KNN) [70], precum și optimizări bazate pe biogeografie (Biogeography-Based Optimization – BBO) și SVM [71]. Comparativ cu metoda bazată pe o fereastră temporală glisantă (Multi-time-scaled Sliding Window based Automatic Emotion Variation Detection - MSW-AEVD) și

HMM [72], metoda propusă oferă rezultate mai bune exceptând cazul în utilizăm SALAS, când arhitectura propusă este mai puțin performantă, deoarece MSW-AEVD + HMM [72] este axată pe emoții induse, iar SALAS conține rostiri colectate în astfel de condiții. Această cercetare a fost publicată în [73].

Pentru **predicția dimensiunilor FFM prin analiză vocală**, cea mai performantă structură neuronală o reprezintă folosirea unui singur FFNN pentru predicția tuturor celor 5 dimensiuni FFM, similar analizei faciale, ceea ce arată că și din punct de vedere vocal cele 5 dimensiuni sunt corelate. Obținem o acuratețe de 76,1% atunci când DSC este utilizat pentru antrenare și testare, în timp ce dacă păstrăm DSC pentru antrenare și testăm folosind DSR, acuratețea scade cu mai puțin de 2%. Cea mai mare acuratețe este obținută pentru „Extroversie” (78,4%), urmată de „Deschidere spre Experiență” (77,2%) și „Agreeabilitate” (76,3%). Timpul maxim necesar predicției este de 120s, mult mai rapid decât chestionarul FFM (21,1 minute), arhitectura putând înlocui chestionarul FFM și putând fi utilizată pentru monitorizare în timp real. Determinăm și în acest caz un set de corelații între caracteristicile vocale și dimensiunile FFM, unice în literatură, care conduc la o creștere a acurateței de 7%. Prin comparație cu alte metode din literatură, arhitectura propusă surclasează metodele bazate pe SVM cu nucleu liniar [74], însă algoritmi bazați pe transformată Wavelet și rețele neuronale convoluționale (Convolutional Neural Networks – CNN) [75] sau predicție liniară în domeniul frecvență (Frequency Domain Linear Prediction - FDLF) prin SVM [76] oferă o acuratețe cu 0,7%, respectiv 3,7% mai mare.

Pentru **predicția temperamentului prin analiză vocală**, structura optimă este obținută când utilizăm FFNN-uri dedicate pentru fiecare temperament, arătând, similar analizei faciale, că tipurile de temperament sunt slab corelate. Cea mai mare acuratețe (79%) este obținută când sistemul este antrenat și testat folosind DSC, în timp ce, dacă menținem DSC pentru antrenare și testăm sistemul folosind DSR, acuratețea scade cu mai puțin de 4%. Astfel, DSC este esențial doar pentru antrenare, în timp ce pentru testare se pot folosi rostiri aleatoare, sugerând faptul că arhitectura propusă este, și aici, una practică. Timpul maxim necesar pentru a prezice cu acuratețe temperamentul este de 78s, mult mai mic decât cel necesar completării chestionarului FIRO-B, ceea ce face acest sistem atractiv pentru a înlocui chestionarul standard, cu avantajul că este mai rapid, mai practic și înlătură subiectivitatea evaluării asociate metodei chestionarului. Obținem cea mai mare acuratețe pentru tipurile de temperament sangvinic și melancolic. Pentru fiecare temperament identificăm, pentru prima dată în literatură, seturi de caracteristici vocale cu care sunt corelate și, modificând arhitectura pentru a cuprinde aceste corelări, acuratețea sistemului crește cu până la 7%. Comparată cu alte metode care analizează vocea cu scopul predicției temperamentului, arhitectura propusă este cea mai performantă din literatură.

Pentru **predicția stărilor emoționale pe baza analizei vocale**, structura optimă este obținută atunci când folosim un singur FFNN pentru a modela toate cele trei stări emoționale, similar analizei faciale. Cea mai mare acuratețe este obținută atunci când DSC este utilizat atât pentru antrenare cât și pentru testare: 86,3% (depresie), 81,6% (anxietate) și 83,3% (stres). Atunci când folosim DSC pentru antrenare și DSR pentru testare obținem o scădere a acurateței de doar 2%. Cea mai mare acuratețe este obținută pentru nivelul *Normal* și pentru nivelul *Sever extrem*, în timp ce pentru nivelurile intermediare acuratețea este mai scăzută și acestea sunt confundate cel mai des cu nivelurile vecine. Timpul maxim necesar predicției este 96s, mult mai rapid decât chestionarul folosit pentru evaluarea nivelurilor DASS, putându-l înlocui dacă predicția nivelurilor intermediare este îmbunătățită. Determinăm, pentru prima dată în literatură, un set de corelații între nivelurile DASS și caracteristicile vocale

analizate care, integrate în arhitectura propusă, duc la o creștere a acurateței cu până la 3%. Comparând arhitectura propusă cu alte metode din literatură, obținem cea mai mare acuratețe pentru predicția stresului, pentru depresie metodele bazate pe ANN și SVM [77] și pe rețele neuronale convoluționale dinamice (Dynamic Convolutional Neural Networks – DCNN) [78] oferind o acuratețe mai mare.

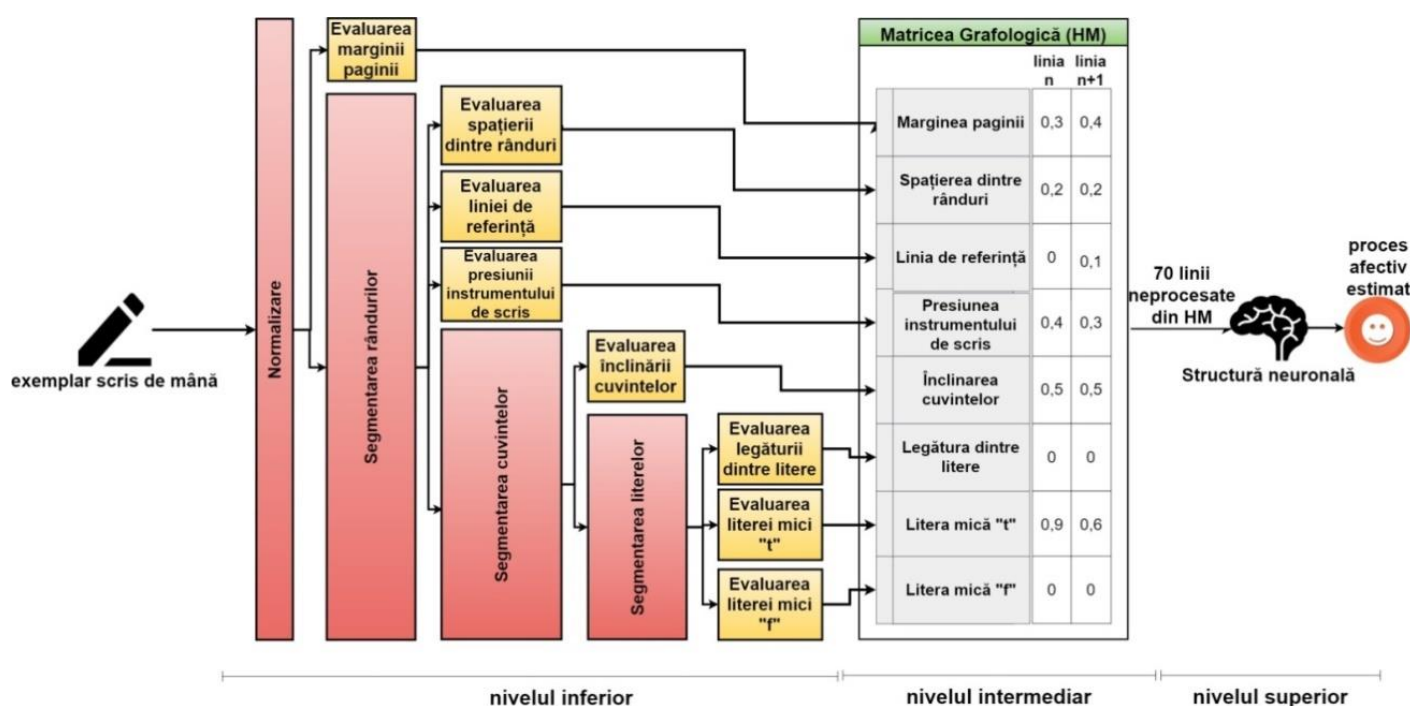
## CAPITOLUL 6 - ANALIZA CARACTERISTICILOR GRAFOLOGICE

Crearea unor sisteme care să fie capabile să recunoască automat procesele afective pe baza scrisului de mână, fără necesitatea unui observator sau evaluator uman, ar putea oferi domeniului grafologiei relevanța necesară pentru a fi studiat cu mai multă încredere [79]. Astfel de sisteme ar fi utile pentru analiză psihologică și punerea diagnosticelor medicale atât în sfera psihologică [80], cât și în cea fiziologică [81]. În cadrul unei analize grafologice standard, grafologii analizează scrisul de mână cu scopul de a determina un set de caracteristici, fiecare conținând anumite informații legate de starea emoțională, trăsăturile de personalitate precum și alte aspecte psihologice ale subiectului [82]. Principalele caracteristici grafologice și cele pe care le utilizăm în această cercetare sunt prezentate în Fig. 6.1 [83], alături de structura nivelurilor inferior și intermediar utilizate pentru arhitecturile bazate pe analiză grafologică.

**Nivelul inferior** are scopul de a converti imaginea scanată conținând exemplarul scris de mână într-un set de caracteristici grafologice identificate pe baza unor scoruri de clasificare. Blocul de **normalizare** presupune reducerea zgomotului pentru care folosim 3 filtre (boolean filter [83] pentru a înlătura fundalul texturizat, un filtru de reducere a lățimii pentru sharpening și un filtru adaptiv de mascare a neclarităților [84] pentru ajustarea contrastului), netezirea conturului pentru a reduce posibilele erori apărute din cauza mișcărilor nedorite ale mâinii subiectului în timpul scrisului pentru care utilizăm metoda de mediere locală ponderată optimă [85], compresia imaginii pentru care utilizăm histograma integrală modificată [86] pentru a reduce timpul de procesare necesar predicției și a permite funcționarea în timp real și izolarea scrisului pentru care utilizăm metoda de reducere a spațiului alb (white space thinning) [87] deoarece este simplă, eficientă și rapidă. Pentru **segmentarea și analiza la nivelul rândurilor** folosim metoda de proiectare a profilurilor verticale (Vertical Projection Profile – VPP) [88]. Caracteristica spațierii dintre rânduri se clasifică determinând cantitatea de pixeli care se suprapun între dreptunghiurile delimitate pentru 2 rânduri consecutive. Caracteristica liniei de referință este determinată pe baza metodei descrise în [89] prin care studiem densitatea pixelilor în fiecare dreptunghi care delimitează un rând segmentat și rotim acest dreptunghi până când cea mai mare densitate de pixeli este centrată orizontal. Caracteristica presiunii instrumentului de scris se determină pe baza metodei de segmentare pe histogramă pe baza tonurilor de gri (grey-level thresholding) [91], analizând valorile din spectrul dreptunghiului care delimitează rândul curent și calculând media pentru rândul segmentat.

Pentru **segmentarea și analiza la nivelul cuvintelor** generăm VPP [88] pentru a determina densitatea pixelilor pentru fiecare coloană verticală și determinăm coloanele cu cea mai scăzută densitate, acestea fiind considerate posibile spații între cuvinte [88]. Pentru caracteristica înclinării literelor folosim tehnica detaliată în [90], calculând densitatea de probabilitate a pixelilor verticali pentru diferite unghiuri și pentru fiecare coloană din histogramă determinând numărul de pixeli și împărțindu-l la cel mai înalt și cel mai jos pixel din segmentul de cuvânt analizat, valorile din aceste coloane fiind adunate și unghiul unde suma calculată este cea mai mare reprezentând înclinarea cuvântului. Pentru **segmentarea și analiza la nivelul literelor** din fiecare cuvânt

delimitat folosim transformata lărimii liniei (Stroke Width Transform - SWT) pentru a determina media lărimii liniilor din acel cuvânt și creăm un VPP pentru segmentul de cuvânt determinând coloanele unde valoarea proiectată este mai mică de 8% (prag determinat prin TAE) din maximum valorii proiectate în cuvânt. Pentru caracteristica legăturii dintre litere folosim algoritmul de segmentare a literelor descris anterior și comparăm lăimea fiecărei linii care conectează 2 dreptunghiuri delimitatoare pentru litere consecutive cu lăimea medie a liniilor din cuvânt. Pentru caracteristicile literă mică “t” și literă mică “f” folosim metoda de potrivire a șabloanelor (Template Matching – TM) pentru a compara fiecare literă cu un set de modele predefinite ale literei “t” din baza de date MNIST [91], măsurând gradul de potrivire prin similaritate euclidiană [92]. Aplicând metodele detaliate anterior, fiecare dintre caracteristicile grafologice analizate este clasificată cu o acuratețe de peste 90%. Nivelul inferior oferă scorurile de clasificare pentru fiecare caracteristică a scrisului de mână sub forma unei valori normalizate în intervalul [0;1]. Pentru fiecare literă din exemplar și fiecare caracteristică asociată ei, scorurile de clasificare sunt furnizate nivelului intermediar și stocate în Matricea Grafologică.



**Fig. 6.1** Structura nivelurilor inferioare și intermediare pentru analiza caracteristicilor grafologice

Pentru **predicția emoției pe baza analizei grafologice**, structura optimă este obținută când utilizăm un singur FFNN pentru a modela toate cele 6 emoții. Din punct de vedere grafologic, emoțiile nu sunt discrete, analiza lor corelată rezultând într-o acuratețe mai ridicată. Cea mai mare acuratețe este obținută când DSC este folosit atât pentru antrenare cât și pentru testare (68,83%), similar arhitecturilor anterioare, observând o scădere de 1,5% a acurateței când DSC este menținut pentru antrenare și DSR este folosit pentru testare, concluzionând că exemplare cu text predefinit sunt necesare doar pentru antrenare, utilizatorii finali putând scrie orice text doresc fără a se observa un impact semnificativ asupra performanței sistemului. Timpul maxim necesar pentru a prezice cu acuratețe emoția subiectului este de 9 minute, mai mic decât timpul necesar pentru a completa DEQ, arhitectura putând înlocui metoda chestionarului dacă acuratețea predicției este îmbunătățită. Obținem cea mai mare acuratețe pentru „tristețe” (72,92%), urmată de „fericire” (71,49%), celelalte emoții oferind o acuratețe mai mică de 70%. Determinăm, pentru prima dată în literatură, un set de caracteristici grafologice asociate cu fiecare dintre cele 6 emoții și, modificând arhitectura pentru a furniza emoția corespunzătoare când aceste caracteristici se regăsesc în combinațiile respective, acuratețea crește cu 6%. Folosim MENTAL pentru a testa alte metode

folosite în literatură, rezultatele arătând că arhitectura propusă oferă rezultate similare. Această arhitectură a fost utilizată pentru predicția tensiunii arteriale pe baza caracteristicilor grafologice, lucrare publicată în [93], atingând o acuratețe a predicției de peste 90%.

Pentru **predicția dimensiunilor FFM pe baza analizei grafologice**, structura optimă o reprezintă utilizarea unui singur FFNN pentru predicția tuturor celor 5 dimensiuni. Cea mai mare acuratețe este obținută când DSC este utilizat atât pentru antrenare cât și pentru testare (82,34%), acuratețea scăzând cu mai puțin de 2% când DSC este utilizat pentru antrenare și DSR pentru testare. Cea mai mare acuratețe este obținută pentru „Deschidere spre Experiență” (84,6%), urmată de „Extroversie” (83,58%) și „Neuroticism” (82,39%). Timpul maxim necesar predicției dimensiunilor FFM este de aproximativ 8 minute, mai rapid decât chestionarul FFM, putându-l înlocui cu succes datorită avantajului de a înlătura subiectivitatea analizei. Identificăm o serie de corelații între cele 5 dimensiuni FFM și caracteristicile grafologice analizate, unice în literatură. Integrând aceste corelații în arhitectura propusă, acuratețea crește cu peste 6%. Utilizăm MENTAL pentru a testa alte metode din literatură și arătăm că arhitectura propusă oferă cele mai performante rezultate din literatură, singura metodă care atinge acuratețe similară fiind cea bazată pe combinația de clasificatori SVM, AdaBoost și KNN [94]. Cercetarea a fost publicată în [95] și extinsă în [96] pentru predicția indicatorului psihologic de tip Myers-Briggs și în [97] prin fuziunea metodei bazate pe FFNN cu SVM, cu o creștere a acurateței de până la 2%.

Pentru **predicția temperamentului pe baza caracteristicilor grafologice**, structura neuronală optimă este aceea în care se utilizează FFNN-uri dedicate pentru fiecare temperament. Din punct de vedere grafologic, temperamentele sunt independente, analiza lor combinată ducând la o scădere a acurateței predicției. Obținem cea mai mare acuratețe când DSC este folosit atât pentru antrenare cât și pentru testare (81,7%). Similar, observăm o scădere a acurateței de doar 4% când DSC este menținut pentru antrenare și DSR este folosit pentru testare. Timpul maxim necesar pentru a prezice cu acuratețe temperamentul este de 7 minute, mai mic decât cel necesar pentru a completa chestionarul FIRO-B. Abordarea noastră este astfel mai rapidă și mai eficientă decât chestionarul standard și sistemul poate fi utilizat ca un înlocuitor atractiv pentru aceasta. Obținem cea mai mare acuratețe pentru tipurile de temperament supin, sanguin și melancolic. Determinăm și în acest caz, pentru prima dată în literatură, un set de corelații între caracteristicile grafologice analizate și tipurile de temperament și, prin utilizarea acestora pentru optimizarea arhitecturii, acuratețea crește cu 6%. Utilizăm MENTAL pentru a testa alte metode și a-i compara cu arhitectura propusă și arătăm că aceasta oferă cea mai mare acuratețe.

Pentru **predicția stărilor emoționale pe baza caracteristicilor grafologice**, optim este să utilizăm un singur FFNN pentru toate cele 3 stări emoționale, acestea fiind puternic corelate din punct de vedere grafologic, confirmând studiile anterioare care identifică o astfel de legătură [98]. Cea mai mare acuratețe este obținută când DSC este utilizat atât pentru faza de antrenare cât și pentru testare: 82,69% (depresie), 79,84% (anxietate) și 81,53% (stres). Dacă folosim DSC doar pentru antrenarea sistemului și îl testăm pe DSR, acuratețea scade cu mai puțin de 2%. Astfel, subiecții pot alege textul pe care îl scriu, sistemul oferind rezultate aproape la fel de precise ca atunci când exemplarele conțin un text predefinit dacă sistemul a fost antrenat anterior pe exemplare controlate de la subiecți multipli. Observăm că cea mai mare acuratețe este obținută pentru nivelurile *Normal* și *Sever extrem*, în timp ce pentru nivelurile intermediare acuratețea este mai scăzută și acestea sunt confundate cel mai des cu nivelurile vecine. Timpul maxim necesar predicției este de 10 minute, mai rapid decât completarea

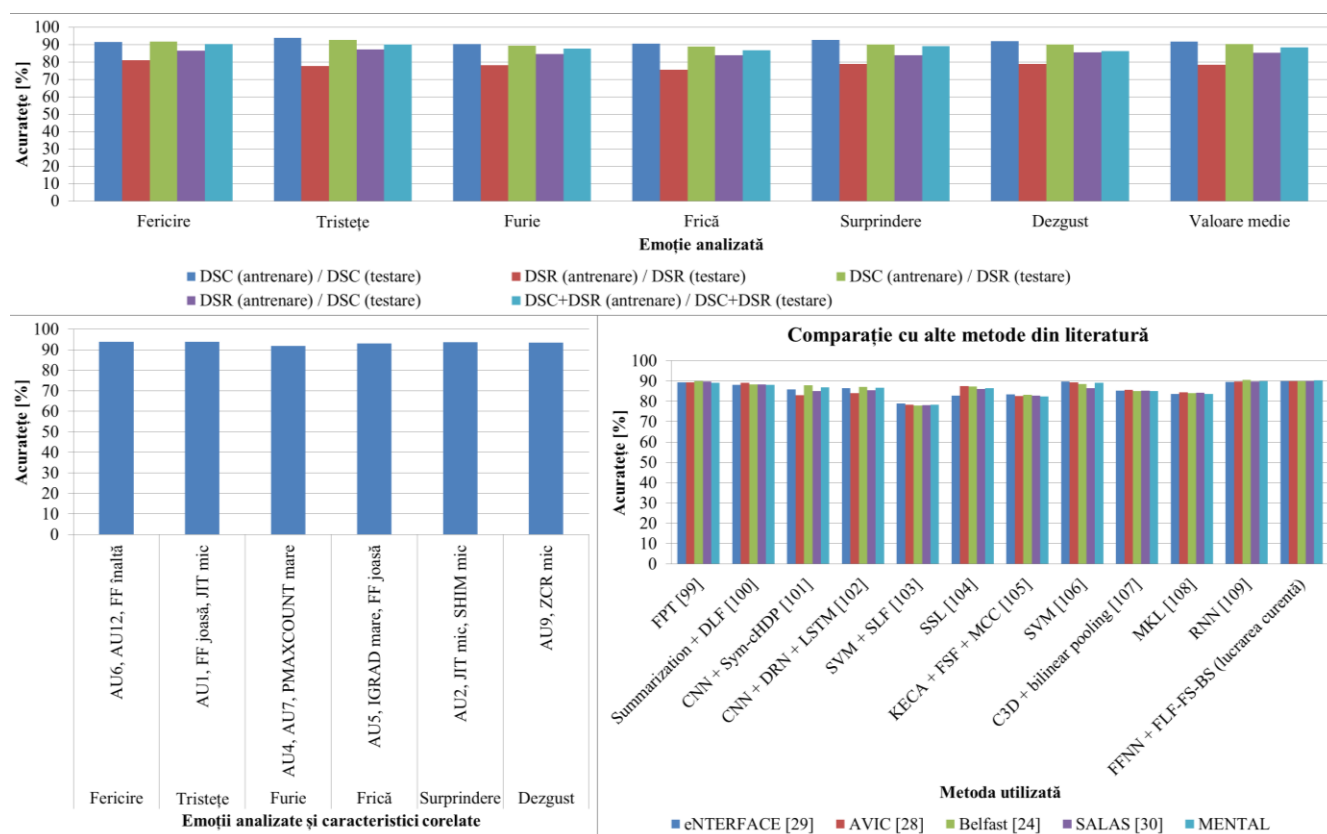
chestionarului (SAQ), oferind posibilitatea monitorizării celor 3 stări emoționale la orice moment de timp fără nevoia de a fi interpretate de un psiholog, de asemenea înlăturând și subiectivitatea specifică chestionarului SAQ. Identificăm, pentru prima dată în literatură, o serie de corelații între caracteristicile grafologice analizate și cele trei stări emoționale. Prin integrarea acestor corelații în arhitectura propusă, acuratețea crește cu până la 3%. Comparăm arhitectura propusă cu metoda bazată pe Random Forest Modeling (RFM) [33] utilizând MENTAL și EMOTHAW. Observăm că arhitectura propusă surclasează metoda bazată pe RFM când este testată pe ambele baze de date, arătând robustețe prin obținerea unor rezultate consistente când este testată pe baze de date diferite.

## CAPITOLUL 7 - ANALIZA MULTIMODALĂ

În capitolele anterioare am evaluat posibilitatea predicției diferitelor procese afective umane pe baza unor analize unimodale, studiind, succesiv, caracteristicile faciale, vocale și grafologice. În acest capitol ne propunem să identificăm modurile optime de fuziune a arhitecturilor unimodale anterioare cu scopul îmbunătățirii predicției proceselor afective pentru ca aceste arhitecturi să fie mai atractive pentru utilizarea în diferite aplicații zilnice și în timp real, precum și pentru a înlocui chestionarele specifice care au dezavantajul de a fi evaluate deseori subiectiv. Deoarece ne dorim o evaluare în timp real, caracteristicile grafologice sunt utilizate doar pentru a ajuta celelalte două tipuri de caracteristici să prezică mai bine procesele afective întrucât achiziționarea și evaluarea scrisului de mână nu se poate realiza concomitent cu a celorlalte caracteristici și nici nu oferă abilitatea de monitorizare în timp real. Întrucât, așa cum am arătat, caracteristicile grafologice oferă o acuratețe ridicată pentru predicția aspectelor legate de personalitate și temperament, folosim un exemplar scris de mână al subiectului analizat cu scopul de a determina un set de parametri comportamentali colectați într-un vector numit Amprentă Comportamentală (Behavioral Stamp - BS), care este utilizat pentru a pondera cele 2 caracteristici (vocală și facială) atunci când se realizează fuziunea acestora. Propunem astfel 5 tipuri de fuziune: **fuziune la nivelul caracteristicilor pentru față și voce** (Feature-level fusion for Face and Speech - FLF-FS) prin utilizarea unui FFNN sau set de FFNN-uri care analizează vectorul combinat al caracteristicilor vocale și faciale, **fuziune la nivelul caracteristicilor pentru față și voce cu amprentă comportamentală** (Feature-Level fusion for Face and Speech with Behavioral Stamp - FLF-FS-BS) în care păstrăm FLF-FS și adăugăm un nivel suplimentar în care rezultatele furnizate de FLF-FS sunt ponderate pe baza unei BS determinate prin antrenarea unui alt FFNN pe exemplare scrise de mână provenind de la aceiași subiecți, **fuziune la nivelul scorurilor pentru față și voce** (Score-Level fusion for Face and Speech - SLF-FS) calculând maximum dintre scorurile furnizate de cele două arhitecturi multimodale (vocală și facială), **fuziune la nivelul scorurilor pentru față și voce cu amprentă comportamentală** (Score-Level fusion for Face and Speech with Behavioral Stamp - SLS-FS-BS) în care modificăm SLF-FS calculând maximum scorurilor ponderate furnizate de cele două arhitecturi unimodale, ponderile fiind determinate pe baza analizei unui exemplar scris de mână care, printr-un FFNN, determină o BS și **fuziune la nivel decizional pentru față și voce** (Decision-Level fusion for Face and Speech - DLF-FS) pentru caracteristicile vocală și facială care este activat doar dacă, de-a lungul a 5s consecutive, rezultatele finale furnizate de ambele arhitecturi unimodale sunt stabile și aceleași. Aceste metode de fuziune bazate pe amprentă comportamentală sunt unice în literatură, fiind proprii acestei teze.

Pentru **predicția emoției prin analiză multimodală**, acuratețea cea mai mare este obținută atunci când utilizăm FLF-FS-BS, componenta grafologică adăugând peste 3% mai multă acuratețe comparativ cu fuziunea

facial-vocală de tip FLF-FS, ceea ce arată că tiparul comportamental determinat pe baza acestei componente adaugă valoare acurateței predicției emoției. Structura optimă o reprezintă utilizarea unui FFNN comun cu un singur nivel ascuns, având funcția de activare *tanh* pentru nivelul ascuns și funcția de activare *sigmoid* pentru nivelul de ieșire. Numărul optim de neuroni ascunși determinat prin TAE este de 980, rata de învățare optimă este 0,2, momentum-ul optim este 0,03 și este nevoie de 25000 de epoci de antrenare pentru a învăța tiparele. Configurația optimă a FFNN-ului utilizat pentru determinarea BS are un singur nivel ascuns, cu funcția de activare *ReLU* pentru nivelul ascuns și funcția *sigmoid* pentru nivelul de ieșire. Numărul optim de neuroni ascunși determinat prin TAE este de 230, rata de învățare optimă este 0,2, momentum-ul optim este 0,04 și este nevoie de 12000 de epoci de antrenare pentru a învăța tiparele.



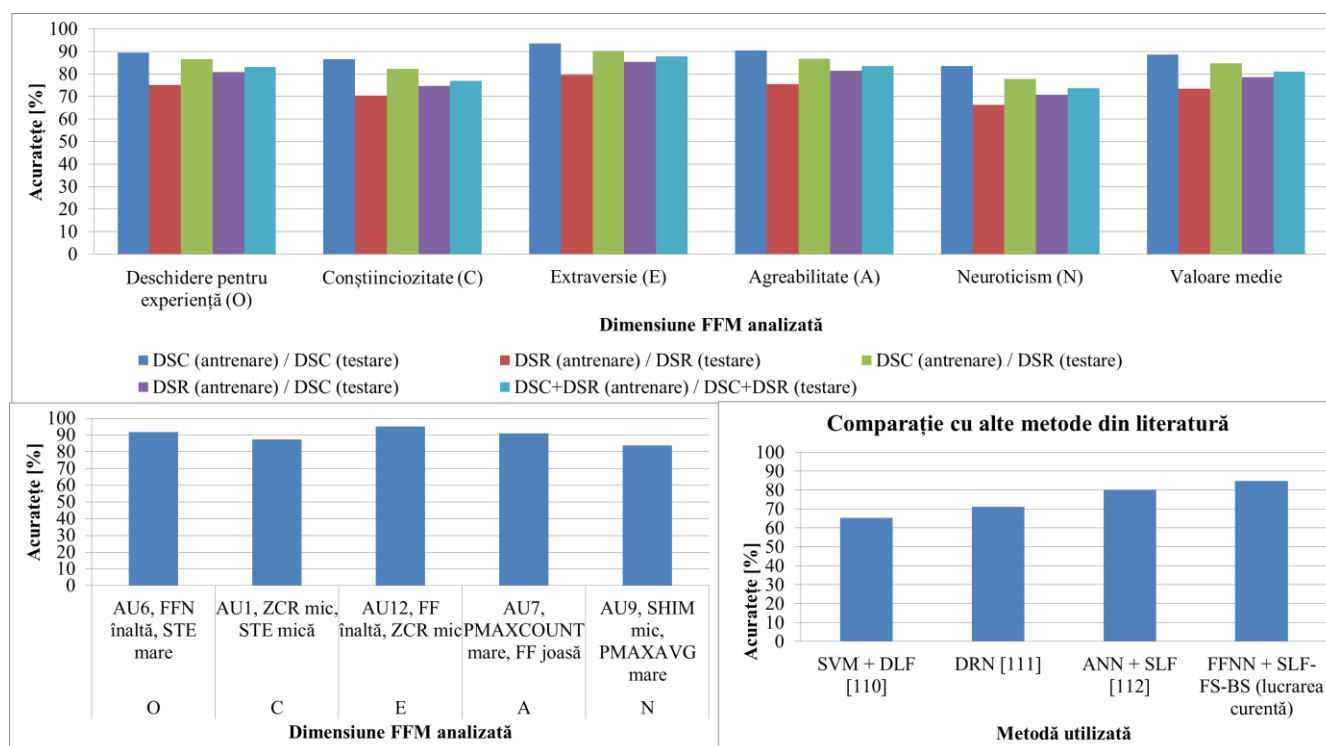
**Fig. 7.1** Predicția emoției prin analiză multimodală

Ca și în cazul arhitecturilor unimodale, cea mai mare acuratețe este obținută când DSC este utilizat atât pentru antrenarea cât și pentru testarea sistemului: 91,84%. Acuratețea scade cu mai puțin de 1,5% atunci când păstrăm DSC pentru antrenarea sistemului și îl testăm folosind DSR. Utilizatorul final nu trebuie să furnizeze decât un exemplar scris de mână pentru a calibra sistemul, urmând ca orice evaluare ulterioară a emoției sale să se poată realiza în condiții aleatoare. Este nevoie de 94s, în medie, pentru a determina emoția subiectului cu acuratețe mare (90,46%), mult mai rapid decât timpul necesar completării DEQ, ceea ce face sistemul pretabil pentru a înlocui chestionarul, oferind posibilitatea monitorizării în timp real a subiectului, precum și o acuratețe ridicată. Obținem cea mai mare acuratețe pentru „tristețe” (92,84%), urmată de „fericire” (91,84%), „surprindere” (90,12%) și „dezgust” (90,03%). Observăm o echilibrare a acurateții obținute pentru fiecare emoție comparativ cu analizele unimodale ceea ce arată că fuziunea de tip FLF-FS-BS reușește să cuprindă caracteristicile relevante emoțional ale fiecărei caracteristici pentru a furniza un rezultat robust și precis. Corelațiile determinate în acest caz sunt similare celor obținute prin analiză unimodală întărind ideea unei arhitecturi robuste. Folosind aceste combinații pentru îmbunătățirea arhitecturii, obținem o acuratețe de peste



91%. Comparăm metodele din literatură cu arhitectura propusă pe mai multe baze de date folosind fuziunea FLF-FS (întrucât celelalte baze de date nu conțin o componentă grafologică) și observăm că metoda noastră oferă acuratețe mai mare decât majoritatea arhitecturilor propuse în literatură.

Pentru **predicția dimensiunilor FFM prin analiză multimodală**, metoda de fuziune optimă este SLF-FS-BS. FFNN-ul optim utilizat pentru generarea BS are 2 niveluri ascunse, funcția de activare *tanh* pentru cele 2 niveluri ascunse și funcția *sigmoid* pentru nivelul de ieșire. Numărul optim de neuroni ascunși determinat prin TAE este de 55 pentru primul nivel ascuns și 35 pentru al doilea nivel ascuns, rata de învățare optimă este 0,2, momentum-ul optim este 0,02 și este nevoie de 12000 de epoci pentru a învăța tiparele. Este nevoie de un bias unitar în al doilea nivel ascuns și antrenarea durează 4 ore.



**Fig. 7.2** Predicția dimensiunilor FFM pe baza analizei multimodale

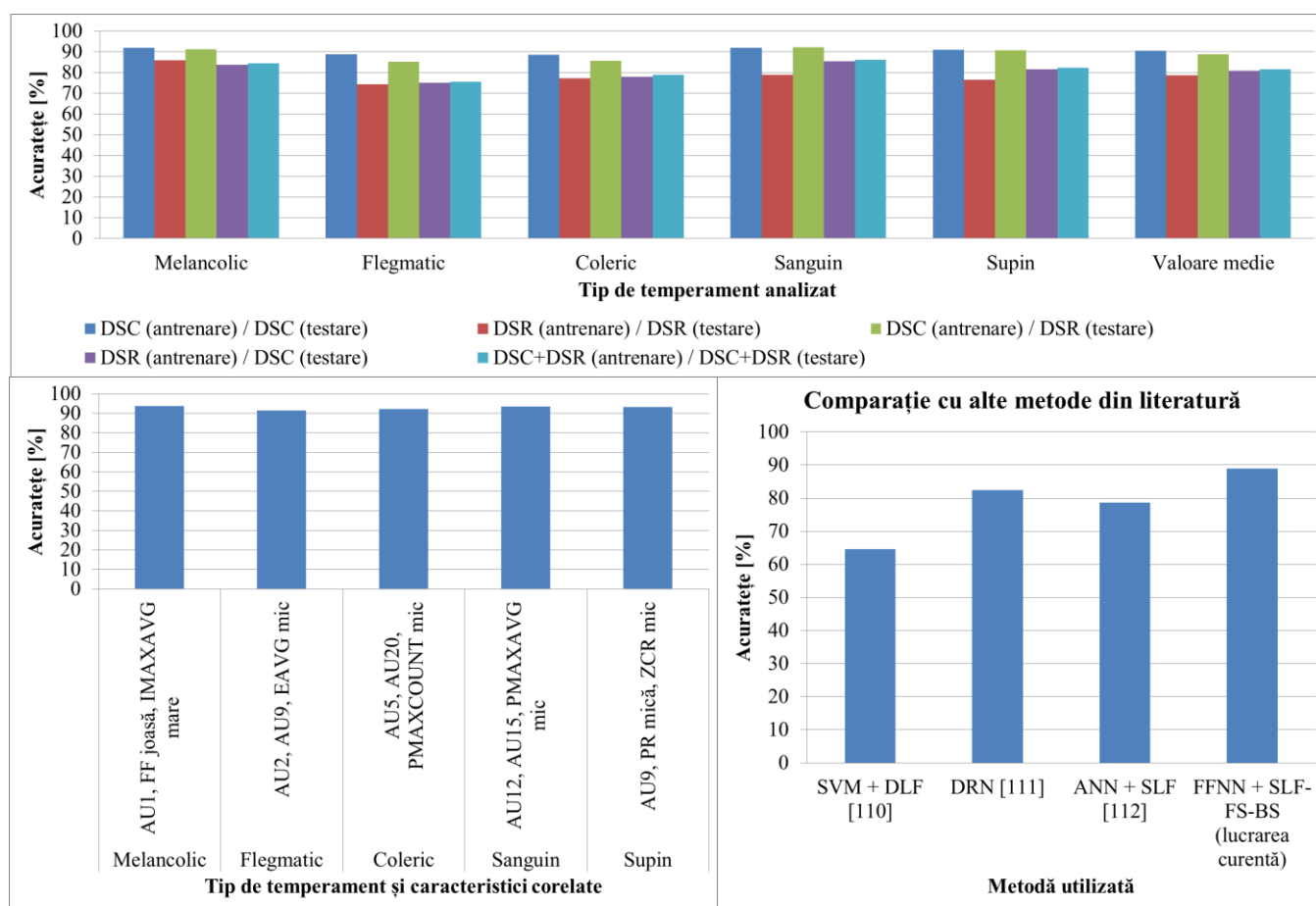
Acuratețea cea mai mare este obținută când DSC este utilizat pentru antrenare și testare (88,7%), în timp ce, utilizând DSC pentru antrenare și DSR pentru testare, acuratețea se reduce cu doar 4%. Durata necesară predicției dimensiunilor FFM este de 115s, considerabil mai scăzută decât cea necesară completării chestionarului FFM (de 21,2 minute), arhitectura fiind rapidă și pretabilă pentru monitorizare în timp real. Cea mai mare acuratețe este obținută pentru „Extroversie” (90,25%), urmată de „Deschidere pentru experiență” (86,51%) și „Agreeabilitate” (86,57%). Determinăm corelații similare celor obținute prin analizele unimodale vocală și facială, arhitectura propusă fiind astfel robustă. Folosind aceste combinații de caracteristici vocale și faciale pentru a optimiza arhitectura propusă, acuratețea sistemului crește cu 5%. Testăm arhitectura propusă în comparație cu alte metode din literatură pe baza de date MENTAL și arătăm că aceasta oferă cea mai mare acuratețe, depășind cu 4% cea mai performantă metodă din literatură bazată pe ANN și fuziune la nivelul scorurilor (Score-level fusion – SLF) [112].

Pentru **predicția temperamentului prin analiză multimodală**, cea mai mare acuratețe este obținută folosind SLF-FS-BS, similar arhitecturii pentru predicția personalității prin modelul FFM. Pentru construirea BS folosim un FFNN cu 2 niveluri ascunse, funcția de activare *sigmoid* pentru primul nivel ascuns și pentru nivelul de ieșire și *tanh* pentru al doilea nivel ascuns. Numărul optim de neuroni ascunși determinat prin TAE este de 65 de



neuroni pentru primul nivel ascuns și 65 de neuroni pentru al doilea nivel ascuns, rata de învățare optimă este 0,3, momentum-ul optim este 0,02 și este nevoie de 8000 de epoci de antrenare.

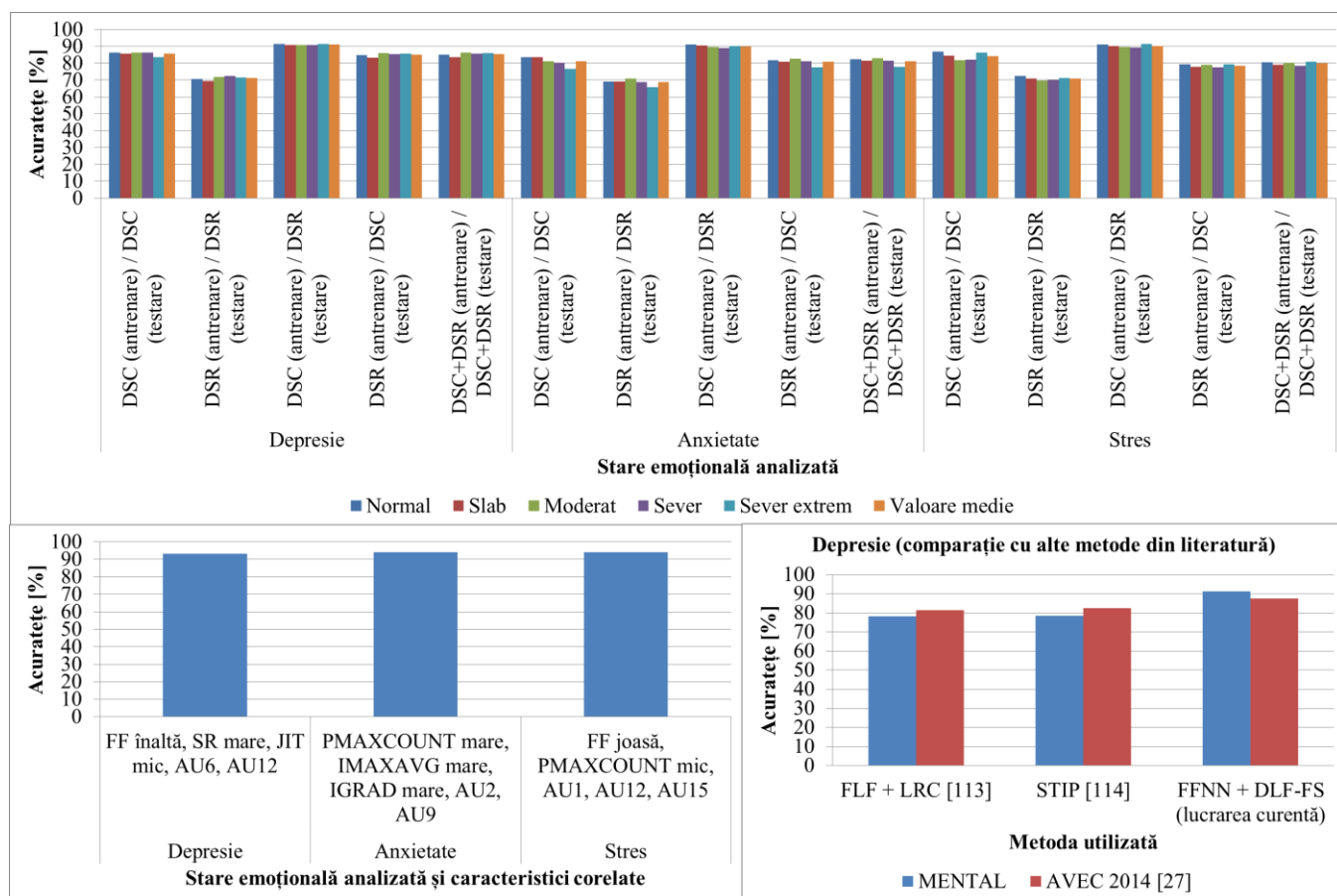
Acuratețea cea mai mare este obținută când DSC este utilizat pentru antrenare și testare (90.41%), aceasta fiind cu doar 1,2% mai mică atunci când DSR este folosit doar pentru testarea sistemului.. Timpul maxim necesar predicției temperamentului este de 117s, considerabil mai mic decât timpul maxim necesar completării chestionarului FIRO-B (14,2 minute), datorită rapidității sale și acurateței ridicate, arhitectura propusă putând fi utilizată cu succes pentru a înlocui metoda chestionarului. Obținem o acuratețe de peste 85% pentru toate cele 5 temperamente, cea mai mare obținându-se pentru temperamentul sanguinic (92,11%), urmat de cel melancolic (91,17%) și de cel supin (90,64%). Corelațiile identificate în acest caz sunt echivalente celor obținute prin analizele unimodale, arătând că arhitectura este robustă. Folosind aceste corelații acuratețea crește cu până la 5%, toate cele 5 temperamente fiind prezise cu o acuratețe mai mare de 91%. Prin comparație cu alte metode din literatură, arhitectura propusă oferă o acuratețe cu peste 9% mai mare decât cea mai performantă metodă din literatură pentru analiza temperamentului prin analiză multimodală.



**Fig. 7.3** Predicția temperamentului pe baza analizei multimodale

Pentru **predicția stărilor emoționale prin analiză multimodală**, metoda de fuziune optimă este DLF-FS. Caracteristicile grafologice nu conțin informație suplimentară față de cea conținută de celelalte 2 caracteristici (faciale și vocale), un sistem optim fiind cel care analizează doar caracteristicile faciale și vocale, fără a mai fi nevoie de analiza grafologică. Cea mai mare acuratețe este obținută când DSC este utilizat atât pentru antrenarea cât și pentru testarea arhitecturii: 91,58% (depresie), 90,2% (anxietate) și 90,48% (stres). Dacă DSC este menținut pentru antrenare și DSR este folosit pentru testare, acuratețea scade cu mai puțin de 1,5%. Toate nivelurile sunt prezise cu acuratețe de peste 89%, ceea ce arată că fuziunea celor 2 caracteristici rezolvă problema întâlnită în cazul analizelor unimodale unde predicția nivelurilor intermediare se realizează cu o acuratețe

considerabil mai mică decât cea a nivelurilor extreme. Analizând caracteristicile din FM și SM prezente la intensitate mare când cele 3 stări emoționale sunt prezise cu acuratețe ridicată, obținem corespondențe similare celor obținute prin analizele unimodale facială și vocală, arătând că arhitectura propusă este una robustă. Folosind aceste combinații de caracteristici vocale și faciale pentru a prezice nivelurile DASS, acuratețea sistemului este îmbunătățită cu până la 4%, toate cele 3 stări emoționale fiind astfel prezise cu peste 93% acuratețe. Realizăm o evaluare comparativă pentru predicția depresiei pe bazele de date MENTAL și AVEC2014 [27] comparând arhitectura propusă cu metoda bazată pe fuziune la nivelul caracteristicilor (Feature-level fusion – FLF) prin clasificatori bazați pe regresie logică (Logical Regression Classifiers - LRC) [113] precum și cea bazată pe caracteristici ale punctelor de interes din spațiu-timp - Space-Temporal Interesting Point – STIP [114]. Rezultatele arată că arhitectura este cea mai performantă metodă din literatură pentru predicția depresiei.



**Fig. 7.4** Predicția stărilor emoționale pe baza analizei multimodale

## CONCLUZII

În această teză am realizat o serie de studii în care am evaluat predicția emoțiilor, personalității, temperamentului și stărilor emoționale atât prin analiză unimodală (expresii faciale, voce sau scris de mână), cât și prin fuziunea acestor trei componente într-o analiză multimodală, determinând cele mai eficiente moduri prin care aceste componente pot fi analizate pentru a asigura cea mai mare acuratețe. Toate arhitecturile propuse sunt unice în literatură, unul dintre scopurile acestei cercetări fiind acela de a determina configurația neuronală ideală pentru fiecare dintre arhitecturile propuse precum și metoda optimă de fuziune, alături de acela de a construi un sistem capabil să recunoască cu acuratețe și într-un interval de timp scurt procesele afective umane, folosind exclusiv metode non-invazive. Fiecare arhitectură este testată folosind metodologiile intra-subiect și inter-subiect și în ambele cazuri se observă că DSC îmbunătățește acuratețea predicției dacă este folosit în faza de antrenare, în timp ce, dacă este folosit în faza de testare, rezultatele obținute nu diferă mult de folosirea DSR în această

fază, ceea ce arată că sistemele propuse sunt practice întrucât subiectul analizat nu trebuie să urmărească niciun videoclip care să inducă emoții sau să scrie vreun text predefinit, sistemul fiind pre-antrenat utilizând înregistrări cu emoție indusă și exemplare cu text predefinit colectate de la alți subiecți. Propunem, de asemenea, două modele de fuziune a componentelor facială și vocală unice în literatură bazate pe un vector BS, construit prin intermediul caracteristicilor grafologice. În cadrul analizei multimodale, construim o arhitectură pentru predicția emoției utilizând o fuziune unică la nivelul caracteristicilor a componentelor vocală și facială cu infuzia componentei grafologice prin intermediul BS, aceasta oferind o acuratețe de 90,46% (cu 3% mai mare decât alte metode din literatură) și o durată medie de predicție de 94s, putând înlocui chestionarul DEQ. Construim, de asemenea, o arhitectură pentru predicția dimensiunilor FFM și temperamentului pe baza unei fuziuni unice la nivelul scorurilor a componentelor vocală și facială cu infuzia componentei grafologice prin intermediul BS, obținând 84,71% acuratețe pentru predicția dimensiunilor FFM (4% mai mare decât alte metode din literatură) cu o durată medie necesară predicției de 115s și 88,85% acuratețe pentru predicția temperamentului (6% mai mare decât alte metode din literatură) cu o durată medie necesară predicției de 117s. Aceste arhitecturi pot înlocui cu succes chestionarele FFM și FIRO-B. În ceea ce privește predicția nivelurilor DASS, prin fuziunea la nivel decizional a componentelor facială și vocală obținem acuratețe de peste 90% pentru toate cele 3 stări emoționale și pentru toate nivelurile de severitate, cu o durată de predicție de 112s, putând înlocui cu succes chestionarul SAQ. Acesta este, totodată, primul studiu din literatură în care se evaluează și corelațiile dintre caracteristicile analizate și procesele afective umane evaluate, determinându-se un set de corespondențe care sunt utilizate pentru a îmbunătăți performanța arhitecturilor, ducând la creșterea cu peste 2% a acurateței predicției emoției, peste 5% pentru predicția dimensiunilor FFM și temperamentului și peste 3% pentru predicția nivelurilor DASS.

Arhitecturile prezentate au o gamă largă de aplicații, pe lângă aceea de a înlocui chestionarele standard. Arhitectura multimodală pentru predicția emoțiilor poate fi integrată cu succes în orice interfață om-computer bazată pe interacțiune facială și vocală, precum și în aplicații pentru consiliere personală sau profesională de tipul psiholog virtual pentru a monitoriza evoluția emoțională a subiectului și a oferi sfaturi personalizate [115]. De asemenea, poate fi utilizată în sisteme de educație mediată de computer, tutorul putând evalua starea emoțională a studenților și putând adapta adecvat conținutul cursului pentru a fi mai bine asimilat de studenți și a îmbunătăți rata de învățare [116]. Sistemul computerizat al unui automobil sau aeronave poate analiza expresiile faciale sau vocea șoferului sau a piloților și lua măsuri preventive pentru a asigura siguranța acestora și a pasagerilor [117]. Întrucât este cunoscut faptul că evoluția emoțională a unui subiect poate furniza informații despre tulburări psihice [118] sau fizice [119] de care acesta suferă, arhitectura propusă poate fi utilizată pentru analiza acestora datorită abilității de a monitoriza emoțiile în timp real. De asemenea, poate ajuta persoanele care suferă de tulburare în spectru autist să se integreze în societate [120]. Bineînțeles, arhitecturile propuse se pot dovedi utile și în departamentele de criminalistică pentru a evalua emoțiile suspectilor și a determina făptașul [59]. Astfel de arhitecturi pot fi utilizate pentru a personaliza un anumit site web sau reclame on-line pe baza emoției subiecților cu scopul de a face informația prezentată mai atractivă pentru aceștia [121]. Nu în ultimul rând, expresiile faciale pot fi utilizate pentru a face sistemele de recunoaștere facială imune la atacuri de tip *spoofing*, așa cum am arătat în [117]. Arhitecturile propuse pentru predicția dimensiunilor FFM pot fi utilizate pentru a determina persoanele aflate sub influența substanțelor [13] precum și a acelor care suferă de boli cronice fizice, cum sunt bolile de

inimă, cancer, diabet sau boli respiratorii, care au fost corelate cu diferite evoluții în timp ale dimensiunilor FFM. Datorită corelării strânse între dimensiunile FFM și tipurile de învățare, arhitecturile propuse pot fi utilizate pentru a adapta conținutul cursurilor sau materialelor educative la stilul de învățare preferat de student. Arhitecturile propuse pentru predicția temperamentului pot fi utilizate pentru construirea de echipe performante, precum și pentru consiliere de carieră și dezvoltare profesională [15]. Arhitecturile propuse pentru determinarea nivelurilor de anxietate, depresie sau stres pot fi utilizate pentru diagnosticarea acuității acestor stări emoționale, având aplicații și în evaluarea psihologică a angajaților [16], învățare adaptivă [17] precum și diagnoza bolilor fizice care au drept simptome anumite evoluții în timp ale acestor stări emoționale [18].

Am observat că o parte dintre arhitecturile prezentate oferă o acuratețe mare care le face pretabile pentru a fi folosite cu succes în aplicațiile amintite anterior, altele însă necesită îmbunătățiri pentru a avea utilitate practică. În studii viitoare se pot considera și altele altor AU-uri alături de cele 31 care sunt analizate în această teză, precum și includerea pozițiilor corpului și gesturilor mâinii care pot purta informații relevante care să permită creșterea acurateței sistemelor (așa cum am arătat în lucrările publicate în [50] și [51]). Similar putem include mai multe caracteristici grafologice sau vocale în analiză pentru a spori acuratețea acestor analize. Contextul în care eșantioanele sunt colectate poate fi, de asemenea, schimbat pentru a îmbunătăți acuratețea, unele dintre procesele afective analizate, cum este „Agreabilitatea”, manifestându-se în contexte sociale care nu sunt create în baza de date propusă. Alte metode bazate pe CNN sau RNN sau alți algoritmi (SVM sau HMM) pot fi evaluate separat sau fuzionate cu metodele din această teză pentru a crește acuratețea și a reduce duratele de procesare.

### LISTA LUCRĂRILOR ORIGINALE

1. M. Gavrilăscu, *Proposed architecture of a fully integrated modular neural network-based automatic facial emotion recognition system based on Facial Action Coding System*, **International Conference on Communications (COMM)**, pp. 1-6, Bucharest, 29-31 May 2014, WOS: 000345844600098.
2. M. Gavrilăscu, *Improved automatic speech recognition system using sparse decomposition by basis pursuit with deep rectifier neural networks and compressed sensing recomposition of speech signals*, **International Conference on Communications (COMM)**, pp. 1-6, Bucharest, 29-31 May 2014, WOS: 000370971100021.
3. M. Gavrilăscu, *Noise robust Automatic speech Recognition system by integrating Robust Principal Component Analysis (RPCA) and Exemplar-based Sparse Representation*, **7th International Conference on Electronics, Computers and Artificial Intelligence**, pp. S-29-S-34, Bucharest, 25-27 June 2015, WOS: 000370971100022.
4. M. Gavrilăscu, *Improved Automatic Speech Recognition System by using compressed sensing signal reconstruction based on L0 and L1 estimation algorithms*, **7th International Conference on Electronics, computers and Artificial Intelligence**, pp. S-23-S-28, Bucharest, 25-27 June 2015, WOS: 000345844600055.
5. M. Gavrilăscu, *Study on determining the Big-Five personality traits of an individual based on facial expressions*, **E-Health and Bioengineering Conference**, pp. 1-6, Iași, 19-21 November 2015, WOS: 000380397900257; *The paper was awarded „Young Researcher” prize.*
6. M. Gavrilăscu, *Study on determining the Myers-Briggs personality type based on individual's handwriting*, **E-Health and Bioengineering Conference**, pp. 1-6, Iași, 19-21 November 2015, WOS: 000380397900256.
7. M. Gavrilăscu, *Recognizing emotions from videos by studying facial expressions, body postures and hand gestures*, **23rd Telecommunications Forum Telfor**, pp. 720-723, Belgrade, 24-26 November 2015, WOS: 000380397000162.
8. M. Gavrilăscu, *Study on using individual differences in facial expressions for a face recognition system immune to spoofing attacks*, **IET Biometrics**, Vol. 5, Issue 3, pp. 236-242, September 2016, IF2016 = 1.382, Q3, WOS: 000382809100010.

9. M. Gavrilescu, *Recognizing human gestures in videos by modeling the mutual context of body position and hands movement*, **Multimedia Systems**, Vol. 23, Issue 3, pp. 381-393, June 2017, IF2017 = 1.703, Q2, WOS: 000438134600003.
10. M. Gavrilescu, N. Vizireanu, *Predicting the Sixteen Personality Factors (16PF) of an individual by analyzing facial features*, **EURASIP Journal on Image and Video Processing**, Vol. 59, August 2017, IF2017 = 1.737, Q2, WOS: 000438134600003.
11. M. Gavrilescu, N. Vizireanu, *Using Off-line handwriting to predict blood pressure level: a neural network-based approach*, **International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures (FABULOUS)**, pp. 124-133, Bucharest, 12-14 October 2017, WOS: 000481658200019.
12. M. Gavrilescu, N. Vizireanu, *Neural Network Based Architecture for Fatigue Detection Based on the Facial Action Coding System*, **International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures (FABULOUS)**, pp. 113-123, Bucharest, 12-14 October 2017, WOS:000481658200018.
13. M. Gavrilescu, *3-Layer Architecture for determining the personality type from handwriting analysis by combining neural networks and support vector machines*, **UPB Scientific Bulletin**, Series C, Vol. 79, Issue 4, pp. 135-150, October 2017, WOS:000424339100011, ISI.
14. M. Gavrilescu, N. Vizireanu, *Predicting the Big Five personality traits from handwriting*, **EURASIP Journal on Image and Video Processing**, Vol. 57, December 2018, IF2018 = 1.534, Q3, WOS: 000438134600003.
15. M. Gavrilescu, N. Vizireanu, *Feedforward Neural Network-Based Architecture for Predicting Emotions from Speech*, **Data**, Vol. 4, Issue 3, pp. 101, July 2019, ISI.
16. M. Gavrilescu, N. Vizireanu, *Predicting Depression, Anxiety, and Stress Levels from Videos Using the Facial Action Coding System*, **Sensors**, Vol. 19, issue 17, August 2019, IF2017 = 3.031, Q1.

**Cele 16 lucrări menționate anterior au fost publicate în 7 reviste ISI (din care 1 Q1, 2 Q2, 2 Q3), 9 conferințe ISI. Lucrările au primit un număr 45 de citări.**

#### BIBLIOGRAFIE

- [1] S. Petrovica, A. Anohina-Naumeca, H. Kemal Ekenel, *Emotion Recognition in affective Tutoring Systems: Collection of Ground-truth Data*, **Procedia Computer Science**, Vol. 104, pp. 437-444, 2017.
- [2] G.H. Graham, J. Unruh, P. Jennings, *The Impact of Nonverbal Communication in Organizations: A Survey of Perceptions*, **International Journal of Business Communication**, 1991.
- [3] M. Szwoch, W. Szwoch, *Emotion Recognition for Affect Aware Video Games*, **Image Processing & Communications Challenges**, Vol. 6, pp. 227-236, 2015.
- [4] A. Fernandez-Caballero et al, *Smart environment architecture for emotion detection and regulation*, **Journal of Biomedical Informatics**, Vol. 64, pp. 55-73, December 2016.
- [5] J.E. Steiner, *Human Facial Expressions in Response to Taste and Smell Simulation*, **Advances in Child Development and Behavior**, Vol. 13, pp. 257-295, 1979.
- [6] M. Pantic, A. Pentland, A. Nijholt, T.S. Huang, *Human Computing and Machine Understanding of Human Behavior: A survey*, **Artificial Intelligence for Human Computing**, pp. 47-71, 2007.
- [7] R.W. Picard, *Affective Computing*, **MIT Press**, 2000.
- [8] M. Clynes, *Sentics: The Touch of the Emotions*, **Anchor Press**, 1978.
- [9] P. Ekman, W.V. Friesen, *Facial Action Coding System: Investigator's Guide*, **Consulting Psychologists Press**, 1978.
- [10] P. Ekman, W.V. Friesen, J.C. Hager, *Facial Action Coding System. Manual and Investigator's Guide*, **Salt Lake City, UT: Research Nexus**, 2002.
- [11] P.T.Jr. Costa, R.R. McCrae, *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual*, **Psychological Assessment Resources**, Odessa, Florida, 1992.
- [12] N. Zilberman, G. Yadid, Y. Efrati, Y. Neumark, Y. Rassovsky, *Personality profiles of substance and behavioral addictions*, **Addictive Behaviors**, Vol. 82, pp. 174-181, July 2018.
- [13] S.J. Karau, R.R. Schmeck, A.A. Avdic, *The big five personality traits, learning styles, and academic achievement*, **Journal on Personality and Individual Differences**, Vol. 51, Issue 4, pp. 472-477, September 2011.
- [14] W.C. Schutz, *FIRO: A Three Dimensional Theory of Interpersonal Behavior*, **Holt, Rinehart, & Winston**, New York, 1958.
- [15] E. Schnell, A. Hammer, *FIRO-B: Technical Guide*, **Mountain View, CPP**, 2000.
- [16] S.H. Lovibond, P.F. Lovibond, *Manual for the Depression Anxiety Stress Scales*, **Psychology Foundation**, Sydney, 1995.
- [17] P. Jafari, F. Nozari, F. Ahrari, Z. Bagheri, *Measurement invariance of the Depression Anxiety Stress Scales-21 across medical student genders*, **International Journal of Medical Education**, Vol. 8, pp. 116-122, 2017.
- [18] M.A.M. Gomma, M.H.A. Elmagd, M.M. Elbadry, R.M.A. Kader, *Depression, Anxiety and Stress Scale in patients with tinnitus and hearing loss*, **European Archives of Oto-Rhino-Laryngology**, Vol. 271, Issue 8, pp. 2177-2184, August 2014.
- [19] Y. Tian, T. Kanade, J. F. Cohn, *Recognizing Action Units for Facial Expression analysis*, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 23, Issue 2, pp. 97-115, February 2001.
- [20] P. Lucey et al., *The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression*, **IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops**, June 2010.

- [21] M.J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, *Coding Facial Expressions with Gabor Wavelets*, **Proceedings of the third IEEE International Conference on Automatic Face and Gesture Recognition**, April 1998.
- [22] M. Pantic, M. Valstar, R. Rademaker, L. Maat, *Web-based database for facial expression analysis*, **IEEE International Conference on Multimedia and Expo**, July 2005.
- [23] A. Dhall, R. Goecke, S. Lucey, T. Gedeon, *Collecting Large, Richly Annotated Facial-Expression Databases from Movies*, **IEEE MultiMedia**, Vol. 19, Issue 3, pp. 34-41, July 2012.
- [24] I. Sneddon, M. McRorie, G. McKeown, J. Hanratty, *The Belfast Induced Natural Emotion Database*, **IEEE Transactions on Affective Computing**, Vol. 3, Issue 1, pp. 32-41, March 2012.
- [25] J. Kossaifi, G. Tzimiropoulos, S. Todorovic, M. Pantic, *AFEW-VA database for valence and arousal estimation in-the-wild*, **Image and Vision Computing**, Vol. 65, pp. 23-36, September 2017.
- [26] N. Sharma, A. Dhall, T. Gedeon, R. Goecke, *Thermal spatiotemporal data for stress recognition*, **EURASIP Journal on Image and Video Processing**, Volume 28, May 2014.
- [27] M. Valstar et al., *AVEC 2014: 3D Dimensional Affect and Depression Recognition Challenge*, **Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge**, pp. 3-10, November 2014.
- [28] B. Schuller, B. Vlasenko, F. Eyben, M. Wollmer, A. Stuhlzatz, A. Wendemuth, G. Rigoll, *Cross-Corpus Acoustic Emotion Recognition: Variances and Strategies*, **IEEE Transactions on Affective Computing**, Vol. 1, Issue 2, pp. 119-131, December 2010.
- [29] O. Martin, I. Kotsia, B. Macq, I. Pitas, *The eNTERFACE' 05 Audio-Visual Emotion Database*, **22nd International Conference on Data Engineering Workshops**, April 2006.
- [30] T. Balomenos et al., *An Introduction to Emotionally Rich Man-Machine Intelligent Systems*, **Eunite**, 2003.
- [31] S. Kim, M. Fillippone, F. Valente, A. Vinciarelli, *Predicting Continuous Conflict Perception with Bayesian Gaussian Processes*, **IEEE Transactions on Affective Computing**, Vol. 5, Issue 2, pp. 187-200, May 2014.
- [32] J.H.L. Hansen, *Analysis and Compensation of Speech under Stress & Noise for Environmental Robustness in Speech Recognition*, **ESCA/NATO Tutorial and Research Workshop on speech Under Stress**, September 1995.
- [33] L. Likforman-Sulem et al., *EMOTHAW: A Novel Database for Emotional State Recognition from Handwriting and Drawing*, **IEEE Transactions on Human-Machine Systems**, Vol. 47, Issue 2, pp. 273-284, April 2017.
- [34] D. France, R. Shiavi, S. Silverman, M. Silverman, D. Wilkes, *Acoustical properties of speech as indicators of depression and suicidal risk*, **IEEE Transactions of Biomedical Engineering**, Vol. 47, Issue 7, pp. 829-837, 2000.
- [35] A. Pampuchidou et al., *Video-based depression detection using local curvelet binary patterns in pairwise orthogonal planes*, **Proc. of annual International Conference of the IEEE Engineering in Medicine and Biology Society**, pp. 3835-3838, August 2016.
- [36] G.B. Huang, *Learning capability and storage capacity of two-hidden-layer feedforward networks*, **IEEE Transactions on Neural Networks**, Vol. 14, Issue 2, pp. 274-281, March 2003.
- [37] J. Li, J.H. Cheng, F. Huang, *Brief Introduction of Back Propagation BP) Neural Network Algorithm and Its Improvement*, **Advances In Computer Science and Information Engineering**, pp. 553-558, 2012.
- [38] S. Masood et al., *Analysis of weight initialization techniques for gradient descent*, **Annual IEEE India Conference**, December 2015.
- [39] S. Kherchaoui, A. Houacine, *Face detection based on a model of the skin color with constraints and template matching*, **International Conference on Machine and Web Intelligence**, October 2010.
- [40] I. Budiman et al., *The effective noise removal techniques and illumination effect in face recognition using Gabor and Non-Negative Matrix Factorization*, **International Conference on Informatics and Computing**, October 2016.
- [41] P. Salembier, *Structuring element adaptation for morphological filters*, **Journal of Visual Communication and Image Representation**, Vol. 3, Issue 2, pp. 115-136, June 1992.
- [42] A.M. Burton, K. Bindemann, *The role of view in human face detection*, **Vision Research**, 49(15), pp. 2026-2036, July 2009.
- [43] G.J. Edwards, C.J. Taylor, T.F. Cootes, *Interpreting face images using active appearance models*, **Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition**, April 1998.
- [44] G. Tzimiropoulos, M. Pantic, *Fast Algorithms for Fitting Active Appearance Modes to Unconstrained Images*, **International Journal of Computer Vision**, Vol. 122, Issue 1, pp. 17-33, March 2017.
- [45] F. Dornaika, J. Ahlberg, *Efficient active appearance model for real-time head and facial feature tracking*, **IEEE International SOI Conference**, October 2003.
- [46] V.P. Kshirsagar, M.R. Baviskar, M.E. Gaikwad, *Face recognition using Eigenfaces*, **3rd International Conference on Computer Research and Development**, March 2011.
- [47] C.T. Liao, Y.K. Wu, S.H. Lai, *Locating facial feature points using support vector machines*, **9th International Workshop on Cellular Neural Networks and Their Applications**, May 2005.
- [48] M. Gavrilescu, *Proposed architecture of a fully integrated modular neural network-based automatic facial emotion recognition system based on Facial Action Coding System*, **International Conference on Communications (COMM)**, pp. 1-6, Bucharest, 2014.
- [49] M. Gavrilescu, N. Vizireanu, *Neural Network Based Architecture for Fatigue Detection Based on the Facial Action Coding System*, **International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures**, pp. 113-123, Bucharest, 12-14 October 2017.
- [50] M. Gavrilescu, *Recognizing emotions from videos by studying facial expressions, body postures and hand gestures*, **23rd Telecommunications Forum Telfor**, pp. 720-723, Belgrade, 24-26 November 2015.
- [51] M. Gavrilescu, *Recognizing human gestures in videos by modeling the mutual context of body position and hands movement*, **Multimedia Systems**, Vol. 23, Issue 3, pp. 381-393, June 2017.
- [52] S. Chin, C.Y. Lee, J. Lee, *An automatic method for motion capture-based exaggeration of facial expressions with personality types*, **Virtual Reality**, Vol. 17, Issue 3, pp. 219-237, September 2013.
- [53] A. D. Setyadi, T. Harsono, S. Wasista, *Human character recognition application based on facial feature using face detection*, **International Electronics Symposium**, September 2015.
- [54] L. Teixeira-Mosquera, J.I. Biel, J.L. Alba-Castro, D. Gatica-Perez, *What your face Vlogs about: Expressions of Emotions and Big-Five Traits Impressions in Youtube*, **IEEE Transactions on Affective Computing**, Vol. 6, Issue 2, pp. 193-205, December 2014.
- [55] M. Gavrilescu, *Study on determining the Big-Five personality traits of an individual based on facial expressions*, **E-Health and Bioengineering Conference**, pp. 1-6, Iași, 19-21 November 2015.

- [56] M. Gavrilescu, N. Vizireanu, *Predicting the Sixteen Personality Factors (16PF) of an individual by analyzing facial features*, **EURASIP Journal on Image and Video Processing**, Vol. 59, August 2017.
- [57] J. Joshi, A. Dhall, R. Goecke, M. Breakspear, G. Parker, *Neural-net classification for spatiotemporal descriptor based depression analysis*, **Proceedings of the 21st International Conference on Pattern Recognition**, November 2012.
- [58] M. Gavrilescu, N. Vizireanu, *Predicting Depression, Anxiety, and Stress Levels from Videos Using the Facial Action Coding System*, **Sensors**, Vol. 19, issue 17, pp. 3693, August 2019.
- [59] W.J. Yoon, K.S. Park, *A study of Emotion Recognition and Its Applications*, **International Conference on Modeling Decisions for Artificial Intelligence**, pp. 455-462, 2007.
- [60] N. Upadhyay, A. Karmakar, *Speech Enhancement using Spectral Subtraction-type Algorithms: A comparison and simulation study*, **Procedia Computer Science**, Vol. 54, pp. 574-588, 2015.
- [61] N.V. Prasad, S. Umesh, *Improved cepstral mean and variance normalization using Bayesian framework*, **IEEE Workshop on Automatic Speech Recognition and Understanding**, December 2013.
- [62] F. Nolan, *Intonational equivalence: An experimental evaluation of pitch scales*, **Proceedings of the 15th International Congress of Phonetic Sciences**, January 2003.
- [63] M. Gavrilescu, *Improved automatic speech recognition system using sparse decomposition by basis pursuit with deep rectifier neural networks and compressed sensing recomposition of speech signals*, **International Conference on Communications (COMM)**, pp. 1-6, Bucharest, 25-27 June 2015.
- [64] M. Gavrilescu, *Noise robust Automatic speech Recognition system by integrating Robust Principal Component Analysis (RPCA) and Exemplar-based Sparse Representation*, **7th International Conference on Electronics, Computers and Artificial Intelligence**, pp. S29-S-35, Bucharest, 25-27 June 2015.
- [65] M. Gavrilescu, *Improved Automatic Speech Recognition System by using compressed sensing signal reconstruction based on L0 and L1 estimation algorithms*, **7th International Conference on Electronics, Computers and Artificial Intelligence**, pp. S-23-S-28, Bucharest, 25-27 June 2015.
- [66] I. Guoth, M. Chmulik, J. Polacky, M. Kuba, *Two-dimensional cepstrum analysis approach in emotion recognition from speech*, **39th International Conference on Telecommunications and Signal Processing**, December 2016.
- [67] H.K. Vydana, P. Vikash, T. Vamsi, K.P. Kumar, A.K. Vuppala, *Detection of emotionally significant regions of speech for emotion recognition*, **Annual IEEE India Conference**, December 2015.
- [68] S.G. Koolagudi, K.S. Rao, *Real life emotion classification using VOP and pitch based spectral features*, **Annual IEEE India Conference**, February 2011.
- [69] L. Fu, C. Wang, Y. Zhang, *Classifier fusion for speech emotion recognition*, **IEEE International Conference on Intelligent Computing and Intelligent Systems**, December 2010.
- [70] M.T. Shami, M.S. Kamel, *Segment-based approach to the recognition of emotions in speech*, **IEEE International Conference on Multimedia and Expo**, October 2005.
- [71] N. Ding, N. Ye, H. Huang, R. Wang, R. Malekian, *Speech emotion features selection based on BBO-SVM*, **10th International Conference on Advanced Computational Intelligence**, June 2018.
- [72] Y. Kan, M. Xu, Z. Wu, L. Cai, *Automatic Emotion Variation Detection in continuous speech*, **Signal and Information Processing Association Annual Summit and Conference**, February 2015.
- [73] M. Gavrilescu, N. Vizireanu, *Feedforward Neural Network-Based Architecture for Predicting Emotions from Speech*, **Data**, Vol. 4, Issue 3, pp. 101, July 2019.
- [74] M. Fallahnezhad, M. Vali, M. Khalili, *Automatic Personality Recognition from reading text speech*, **Iranian Conference on Electrical Engineering**, July 2017.
- [75] M.H. Su et al., *Personality trait perception from speech signals using multiresolution analysis and convolutional neural networks*, **Asia-Pacific Signal and Information Processing Association Annual Summit and Conference**, February 2018.
- [76] S. Jothilakshmi, R. Brindha, *Speaker trait prediction for automatic personality perception using frequency domain linear prediction features*, **International Conference on Wireless Communications, Signal Processing and Networking**, September 2016.
- [77] V. Mitra, A. Tsiartas, E. Shirberg, *Noise and reverberation effects on depression detection from speech*, **IEEE International Conference on Acoustics, Speech and Signal Processing**, May 2016.
- [78] L. He, C. Cao, *Automatic depression analysis using convolutional neural networks from speech*, **Journal of Biomedical Informatics**, Vol 83, pp. 103-111, 2018.
- [79] L. Guarnera, G.M. Farinella, A. Furnari, A. Salici, C. Ciampini, V. Mantraga, S. Battiato, *GRAPHJ: A forensic tool for handwriting analysis*, **19th International Conference on Image Analysis and Processing**, September 2017.
- [80] M.M. Prunty, A.L. Barnett, K. Wilmut, M.S. Plumb, *An examination of writing pauses in the handwriting of children with developmental coordination disorder*, **Research in Developmental Disabilities**, Vol. 35, Issue 11, pp. 2894-2905, November 2014.
- [81] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, A.S. di Freca, *Handwriting analysis to support neurodegenerative diseases diagnosis: A review*, **Pattern Recognition Letters**, May 2018.
- [82] R.N. Morris, *Forensic Handwriting Identification: Fundamental Concepts and Principles*, 2000.
- [83] W.L. Lee, K.C. Fan, *Document Image preprocessing based on optimal Boolean filters*, **Signal Processing**, 80(1), pp. 45-50, 2000.
- [84] S.C.F. Lin et al., *Intensity and edge-based adaptive unsharp masking filter for color image enhancement*, **Optik - International Journal for Light and Electron Optics**, Vol. 127, Issue 1, pp. 407-414, January 2016.
- [85] R. Legault, C.Y. Suen, *Optimal local weighted averaging methods in contour smoothing*, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 18, pp. 690-706, July 1997.
- [86] Y. Solihin, C.G. Leedham, *Integral ratio: a new class of global thresholding techniques for handwriting images*, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 21, pp. 761-768, August 1999.
- [87] K. Chen, F. Yin, C.L. Liu, *Hybrid page segmentation with efficient whitespace rectangles extraction and grouping*, **12th International Conference on Document Analysis and Recognition**, pp. 958-962, 2013.
- [88] V. Papavassiliou, T. Stafylakis, V. Katsouro, G. Carayannis, *Handwritten document image segmentation into text lines and words*, **Pattern Recognition**, Vol. 43, pp. 369-377, 2010.



- [89] M.B. Menhaj, F. Razazi, *A new fuzzy character segmentation algorithm for Persian/Arabic typed texts*, **International Conference on Computational Intelligence**, pp. 151-158, 1999.
- [90] R. Coll, A. Fornes, J. Lladós, *Graphological analysis of handwritten text documents for human resources recruitment*, **12th International Conference on Document Analysis and Recognition**, pp. 1081-1085, July 2009.
- [91] L. Deng, *The MNIST database of handwritten digit images for machine learning research*, **IEEE Signal Processing Magazine**, Vol. 29, Issue 6, pp. 141-142, November 2012.
- [92] O. Golubitsky, S.M. Watt, *Distance-based classification of handwritten symbols*, **International Journal on Document Analysis and Recognition**, Vol. 13, Issue 2, pp. 133-146, June 2010.
- [93] M. Gavrilescu, N. Vizireanu, *Using Off-line handwriting to predict blood pressure level: a neural network-based approach*, **International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures**, pp. 124-133, Bucharest, 12-14 October 2017.
- [94] Z. Chen, T. Lin, *Automatic personality identification using writing behaviors: an exploratory study*, **Behavior and Information Technology**, Vol. 36, Issue 8, pp. 839-845, 2017.
- [95] M. Gavrilescu, N. Vizireanu, *Predicting the Big Five personality traits from handwriting*, **EURASIP Journal on Image and Video Processing**, Vol. 57, December 2018.
- [96] M. Gavrilescu, *Study on determining the Myers-Briggs personality type based on individual's handwriting*, **E-Health and Bioengineering Conference**, pp. 1-6, Iași, 19-21 November 2015.
- [97] M. Gavrilescu, *3-Layer Architecture for determining the personality type from handwriting analysis by combining neural networks and support vector machines*, **UPB Scientific Bulletin, Series C**, Vol. 79, Issue 4, pp. 135-150, 2017.
- [98] E.S. Friedman, D.B. Clark, S. Gershon, *Stress, anxiety, and depression: Review of biological, diagnostic, and nosologic issues*, **Journal of Anxiety Disorders**, Vol. 6, Issue 4, pp. 337-363, December 1992.
- [99] A.S. Patwardhan, *Multimodal mixed emotion detection*, **2nd International Conference on Communication and Electronics Systems (ICCES)**, October 2017.
- [100] S. Zhalehpour, Z. Akhtar, C.E. Erdem, *Multimodal emotion recognition with automatic peak frame selection*, **IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA) Proceedings**, June 2014.
- [101] J. Xue, Z. Luo, K. Eguchi, T. Takiguchi, T. Omoto, *A Bayesian nonparametric multimodal data modeling framework for video emotion recognition*, **IEEE International Conference on Multimedia and Expo**, July 2017.
- [102] P. Tzirakis, G. Trigeorgis, M.A. Nicolau, B.W. Schuller, S. Zafeiriou, *End-to-End multimodal emotion recognition using deep neural networks*, **IEEE Journal of Selected Topics in Signal Processing**, Vol. 11, Issue 8, pp. 1301-1309, October 2017.
- [103] R. Gajsek, V. Struc, F. Mihelic, *Multimodal emotion recognition using canonical correlations and acoustic features*, **20th International Conference on Pattern Recognition**, August 2010.
- [104] Z. Zhang, F. Ringeval, B. Dong, E. Coutinho, E. Marchi, B. Schuller, *Enhanced semi-supervised learning for multimodal emotion recognition*, **IEEE International Conference on Acoustics, Speech and Signal Processing**, May 2016.
- [105] Z. Xie, Y. Tie, L. Guan, *A new audiovisual emotion recognition system using entropy-estimation-based multimodal information fusion*, **IEEE International Symposium on Circuits and Systems**, May 2015.
- [106] A. Agrawal, N.K. Mishra, *Fusion Based Emotion Recognition System*, **International Conference on Computational Science and Computational Intelligence**, December 2016.
- [107] D. Nguyen, K. Nguyen, S. Sridharan, D. Dean, C. Fookes, *Deep spatiotemporal fusion with compact bilinear pooling for multimodal emotion recognition*, **Computer Vision and Image Understanding**, July 2018.
- [108] S. Poria, H. Peng, A. Hussain, N. Howard, E. Cambria, *Ensemble Application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis*, **Neurocomputing**, Vol. 261, pp. 217-230, October 2017.
- [109] G. Cariadakis, K. Karpouzis, M. Wallace, L. Kessous, N. Amir, *Multimodal user's affective state analysis in naturalistic interaction*, **Journal on Multimodal User Interfaces**, Vol. 3, Issue 1-2, pp. 49-66, March 2010.
- [110] L. Batrinca, N. Mana, B. Lepri, F. Pianesi, *Multimodal Personality Recognition in Collaborative Goal-Oriented Tasks*, **IEEE Transactions on Multimedia**, Vol. 18, Issue 4, pp. 659-673.
- [111] Y. Guckuturk et al., *Visualizing apparent personality analysis with deep residual networks*, **IEEE International Conference on Computer Vision Workshops**, October 2017.
- [112] J. Gorbova, I. Lusi, A. Litvin, G. Anbarjafari, *Automated Screening of Job Candidate Based on Multimodal Video Processing*, **IEEE Conference on Computer Vision and Pattern Recognition Workshops**, July 2017.
- [113] H. Dibeklioglu, Z. Hammal, J.F. Cohn, *Dynamic Multimodal Measurement of Depression Severity using deep autoencoding*, **IEEE Journal of Biomedical and Health Informatics**, Vol. 22, Issue 2, pp. 525-536, March 2018.
- [114] L. He, D. Jiang, H. Sahli, *Multimodal depression recognition with dynamic visual and audio cues*, **International Conference on Affective Computing and Intelligent Interaction**, December 2015.
- [115] G. Cariadakis, K. Karpouzis, M. Wallace, L. Kessous, N. Amir, *Multimodal user's affective state analysis in naturalistic interaction*, **Journal on Multimodal User Interfaces**, Vol. 3, Issue 1-2, pp. 49-66, March 2010.
- [116] A. Landowska, *Affective Computing and affective learning - methods, tools and prospects*, **EduAkcja. Magazyn edukacji elektronicznej**, Vol. 1, Issue 5, pp. 16-31, 2013.
- [117] M. Gavrilescu, *Study on using individual differences in facial expressions for a face recognition system immune to spoofing attacks*, **IET Biometrics**, Vol. 5, Issue 3, pp. 236-242, September 2016.
- [118] J.M. Girard, J. F. Cohn, M. H. Mahoor, S. Mavadati, D. P. Rosenwald, *Social Risk and Depression: Evidence from Manual and Automatic Facial Expression analysis*, **Proc. of Int. Conference on Automatic Face and Gesture Recognition**, pp. 1-8, 2013.
- [119] E. L. Rosenberg et al., *Linkages between facial expressions of anger and transient myocardial ischemia in men with coronary artery disease*, **Emotion**, Vol. 1, Issue 2, pp. 07-115, June 2001.
- [120] T. Guha et al, *On quantifying facial expression-related atypicality of children with Autism Spectrum Disorder*, **Proceedings of IEEE International Conference on Acoustics and Speech Signal Processing**, pp. 803-807, 2015.
- [121] A. Kolakowska, a. Landowska, M. Szwoch, M. Wrobel, *Emotion Recognition and Its Applications*, **Advances in Intelligent Systems and Computing**, Vol. 300, pp. 51-62, July 2014.