Universitatea
Transilvania
din Brașov

HABILITATION THESIS

Title: AFFECTIVE COMPUTING APPLIED TO VIRTUAL REALITY BASED-PHOBIA TREATMENT AND LEARNING PROCESS

Domain: Computer Science

Author: Assoc. Prof. Dr. Gabriela MOISE

University: Petroleum-Gas University of Ploiești

BRAŞOV, 2025

CONTENTS

Acknowledgement

First of all, I want to thank my family for the all the support and encouraging me throughout writing this thesis. Also, I want to thank all my collaborators with whom I did research and wrote articles over the years. This allowed me to further my research interests and achieve worthwhile results in my field. Lastly, I want to thank my colleagues from my university and from other universities who supported me during my habilitation process.

(A)Rezumat

În acest rezumat este prezentată pe scurt o selecție a principalelor contribuții științifice ale autoarei acestei teze de abilitare.

În **Capitolul 1** cu titlul "O Privire asupra Calcului Afectiv" este realizată o scurtă introducere în domeniul inițiat de Prof. Picard (1995) [1]. Pe scurt, domeniul calculului afectiv se ocupă cu dezvoltarea tehnologiilor care sunt conștiente de emoțiile umane și are aplicații în sănătate, jocuri, realitate virtuală, sisteme de recomandare, marketing, interfețe om-mașină, etc. Învățarea afectivă este un subdomeniu al calculului afectiv care consideră includerea tehnologiilor-conștiente de emoții în procesul de învățare [2]. Tehnologiile de învățare conștiente de emoțiile oamenilor sunt acele tehnologii care recunosc și manifestă emoții în situații de învățare. Mello&Graesser (2015) au studiat aceste tehnologii și au accentuat beneficiul utilizării lor în procesul de învățare [3]. Tododată, accentuăm îngrijorările legate de utilizarea calculului afectiv și anume, cele legate de riscurile etice. Aplicațiile care au integrate module de recunoașterea emoțiilor cu ajutorul inteligenței artificiale aparțin sistemelor de clasă de risc mare conform reglementării AI Act (2024) și au cerințe specifice [4].

Calculul afectiv cuprinde două subiecte majore: recunoașterea/detectarea emoțiilor și exprimarea emoțiilor de către mașini. Cercetările realizate de autoarea acestei teze includ recunoașterea emoțiilor în contextul tratării fobiilor utilizând aplicații de realitate virtuală și recunoașterea emoțiilor în contextul procesului de învățare.

În **Capitolul 2** cu titlul "Date pentru recunoașterea automată a emoțiilor" sunt prezentate pe scurt modele ale emoțiilor și tipurile de date care pot fi utilizate pentru recunoașterea acestora. Sunt prezentate modelul discret, modelul dimensional și modelul componențial ale emoțiilor, primele două fiind utilizate în modelele din această teză. Pentru cercetările menționate, am utilizat trei seturi de date ce conțin date biofizice. Două seturi de date reprezintă contribuția noastră și au fost utilizate în cercetări publicate de autoarea acestei teze în calitate de coautor în [5, 6, 7, 8, 9, 10, 11, 12].

Setul de date 1 conține semnale EEG (Electroencephalography), EDA (Electrodermal activity) și HR (Heart Rate) achiziționate de la 4 subiecți atât în mediu virtual cât și în mediul real. Pentru aceasta, am utilizat un dispozitiv Acticap Xpress Bundle cu 16 electrozi uscați pentru achiziționarea semnalelor de la canalele: FP1, FP2, FC5, FC1, FC2, FC6, T7, C3, C4, T8, P3, P1, P2, P4, O1 și O2. Semnalele EDA și HR au fost achizționate cu ajutorul unității GSR a unui dispozitiv Shimmers Multi-Sensory.

Setul de date 2 conține semnalele GSR (Galvanic Skin Response), HR și RR (Respiration Rate) achiziționate de la 5 subiecți în două situații. În prima, am măsurat GSR, HR și RR ca referință în timp ce subiecții efectuau următoarele acțiuni: respirație adâncă, mișcări ale capului spre stânga, dreapta, sus, jos, clic cu mâna dreapta pe controlerul HTC Vive și ridicarea mâinii drepte. În cea de a doua situație am efectuat aceleași măsurători în timp ce subiecții au jucat un joc video cu realitate virtuală. Pentru măsurarea HR și GSR am utilizat dispozitivul Shimmer3 GSR+ Unit (https://www.shimmersensing.com/product/shimmer3-gsr-unit/). RR a fost calculată pe baza distanței dintre două dispozitive tracker HTC Vive.

Setul de date 3 este un set de referință și anume setul de date DEAP (Database for Emotion Analysis using Physiological signals) prezentat în [13].

Capitolul 3 cu titlul "Sisteme conștiente de emoții cu realitate virtuală de terapie prin expunere pentru tratatarea fobiilor" este bazat pe articolele publicate de autoarea acestei teze în calitate de coautor în [5], [6], [7], [8], [9], [10], [11], [12], [14], [15], [16]. În prima parte a capitolului este prezentată importanța subiectului Terapie prin Expunere cu Realitate Virtuală (TERV) în tratarea fobiilor. Studiile noastre au fost concentrate pe tratarea acrofobiei, care are o incidență mare afectând 1 din 20 de indivizi conform studiului prezentat în [17]. Utilizarea tehnologiile de RV în tratarea fobiilor a fost pentru prima dată remarcată la sfârșitul anilor 1990 [18], de atunci mulți oameni preferând acest tip de terapie.

Subcapitolul "Modele de recunoaștere automată a emoțiilor pe baza setului de date DEAP" prezintă abordările noastre în ceea ce privește dezvoltarea modelelor de Recunoaștere Automată a Emoțiilor (RAE) utilizând setul de date DEAP.

În primul set de modele am recunoscut sentimentul de frică utilizând două paradigme pentru măsurarea intensității fricii: scala cu 2 nivele (absență/prezență) și scala cu 4 nivele (nu/mic/mediu/mare) [14]. Am construit cinci seturi de intrări pentru fiecare paradigmă: Raw, Power Spectral Density, Petrosian Fractal Dimension, Higuchi Fractal Dimension, Approximate Entropy pentru semnalele EEG și înregistrările fiziologice (hEOG, vEOG, zEMG, tEMG, GSR, Respiration, PPG și temperatură). Au fost testate mai multe modele de învățare automată: patru modele de rețele neuronale adânci și Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Random Forest (RF), k-Nearest Neighbors (kNN) cu mai multe tehnici de selecție a atributelor - Principal Component Analysis (PCA), Sequential Feature Selector (SFS), Fisher selection. Modelul cel mai performant a fost clasificatorul RF fără selecția atributelor cu acuratețea – 93.13%, scorul F1– 93.11% în cazul scalei cu 2 nivele și acuratețea – 85.74%, scorul F1 score – 85.33% în cazul scalei cu 4 nivele. Performanțele au fost obținute pe un set de intrare generat din setul de date DEAP la care am aplicat Power

Spectral Density (PSD) pentru toate cele 32 de canale EEG, în domeniul frecvențelor alpha, beta și theta și am calculat valori medii pentru alpha, beta și theta PSD în zone ale creierului: prefrontal (FP), AF (între FP și F), frontal (F), FC (între F și C), central (C), temporal (T), P (parietal), CP (între C și P), O (occipital) și PO (între P și O). De asemenea, setul de intrare a conținut 8 atribute fiziologice din DEAP, hEOG, vEOG, zEMG, tEMG, GSR, Respirația, PPG și temparatura. Abordarea din [14] a fost extinsă la recunoașterea celor 6 emoții discrete definite de Ekman în [19]. Rezultatele sunt prezentate în [16].

În [15], am luat în considerare recunoașterea fricii numai din semnale GSR și HRV (Heart Rate Variability) extrase din setul de date DEAP și am propus un protocol pentru extragerea atributelor. Protocolul a constat în segmentarea fiecărei înregistrări în două moduri: în trei ferestre fără-suprapuneri și în cinci ferestre cu-suprapuneri. Astfel, am extras 33 tipuri de atribute pentru EDA și 7 tipuri de atribute pentru HRV, obținând în total 40 de tipuri de atribute.

Desfășurarea procesului utilizat în [15] este evidențiată în Figura 1.

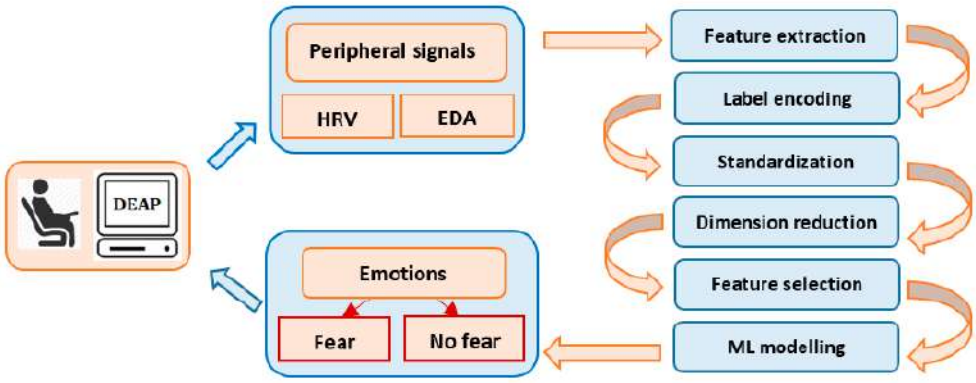

Figura 1. Desfășurarea procesului pentru recunoașterea automată a emoțiilor [15]

Abordarea cu ferestre fără-suprapuneri a constat în divizarea fiecărui segment de 60 de secunde în trei ferestre fără-suprapuneri de câte 20 de secunde fiecare. În abordarea cu ferestre cu-suprapuneri, fiecare segment a fost divizat în cinci ferestre de câte 20 de secunde fiecare și 10 secunde suprapunându-se cu alte segmente.

Pentru ambele abordări, cele mai bune performanțe, peste 89%, au fost atinse cu algoritmii SVM (Support Vector Machine) și GBT (Gradient Boosting Tree). În ceea ce privește scorul ROC AUC, au fost obținute cele mai bune rezultate astfel:

- Pentru setul de date fără-suprapuneri: reducere PCA + SVM – 93.5%.
- Pentru setul de date cu-suprapuneri: GBT – 91.7%.

În subcapitolul "Dezvoltarea de sisteme adaptive TERV cu recunoașterea emoțiilor" este prezentată abordarea noastră pentru tratarea fobiilor utilizând jocuri video cu realitate virtuală.

În primul joc cu RV, jucătorii au fost expuși la diferite niveluri de înălțime din joc în conformitate cu datele fiziologice achiziționate de la aceștia [6], [9], [14]. Arhitectura sistemului TERV propus este prezentată în Figura 2 [6], [9], [14]. Am proiectat și antrenat două rețele neurale profunde (DNN), care au fost integrate în joc: prima rețea estimează intensitatea fricii jucătorilor, cea de a doua rețea determină nivelul de joc care se va afișa utilizatorului. În acest mod, utilizatorii sunt expuși gradual în timp real la diferite niveluri de înălțime adaptate gradului de intensitate a sentimentului de frică.

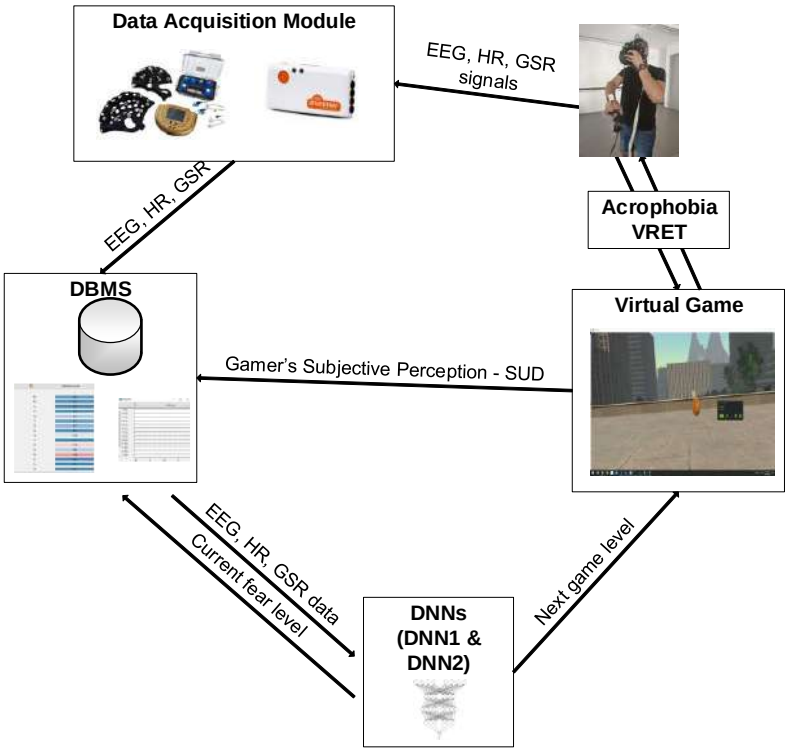

Figura 2. Arhitectura sistemul TERV conștient de emoțiile umane [9]

Sistemul conține un modul pentru achiziția semnalelor EEG, HR și GSR. Semnalele sunt preprocesate și transferate într-un sistem de baze de date, DBMS. Datele extrase din semnalele EEG, HR și GSR reprezintă intrările unui clasificator DNN1 pentru a determina nivelul curent de frică. Utilizând aceleași date și nivelul de frică dorit, cea de a doua rețea DNN2 generează predicția nivelului de joc potrivit unui anumit utilizator.

Scenariul de lucru al jocului este prezentat în Figura 3. Utilizatorul începe jocul cu nivelul l_0 , astfel nivelul curent de joc l_{cr} este l_0. Datele EEG, HR și GSR sunt înregistrate în timpul jocului. Când un utilizator termină un nivel de joc, datele achiziționate sunt transmise primei rețele DNN1, care face predicții asupra nivelului curent de frică. Nivelul următor dorit de frică

se calculează în conformitate cu una din cele două paradigme: paradigma cu scala cu 2 nivele sau paradigma cu scala cu 4 nivele.

În cazul scalei cu 2 nivele formula de calcul pentru nivelul dorit de frică este:

if $fl_{cr} == 0$ then $fl_d = 0$

if $fl_{cr} == 1$ then $fl_d = 1$

În cazul scalei cu 4 nivele formula de calcul pentru nivelul dorit de frică este:

if $fl_{cr} == 0$ or $fl_{cr} == 1$ then $fl_d = fl_{cr} + 1$

if $fl_{cr} == 2$ then $fl_d = fl_{cr}$

if $fl_{cr} == 3$ then $fl_d = fl_{cr} - 1$

Nivelul dorit de frică și datele biofizice sunt intrări pentru cea de a doua rețea neurală profundă, DNN2, care generează următorul nivelul de joc (l_{pr}). Utilizatorii joacă nivelul de joc și datele biofizice sunt înregistrate. Algoritmul se termină când este atins un anumit număr de epoci de joc.

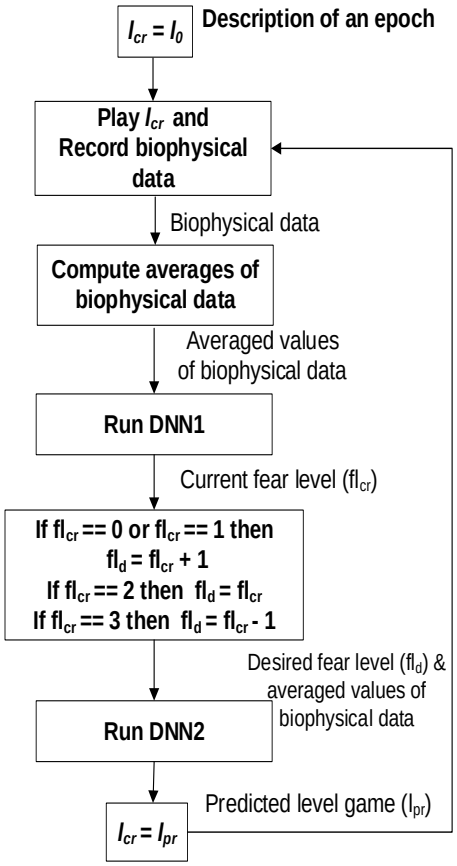

Figura 3. Algoritmul jocului TERV în cazul scalei cu 4 nivele [9]

Am testat trei tipuri de scale pentru evaluarea fricii, scale cu 2, 4 și 11 nivele, obținând următoarele performanțe pentru clasificarea fricii, în termeni de acuratețe:

- Scala cu 2 nivele - 95.51% utilizând un DNN cu 3 straturi ascunse – 300 de neuroni pe fiecare strat, activare RELU, stratul de ieșire – activare Sigmoid, funcția de pierdere Binary cross-entropy.
- Scala cu 4 nivele - 90.49% utilizând un DNN cu 3 straturi ascunse – 300 de neuroni pe fiecare strat, activare RELU, stratul de ieșire – funcția de activare Softmax, funcția de pierdere Logarithmical categorical cross-entropy.
- Scala cu 11 nivele – 85.09% utilizând un DNN cu 3 straturi ascunse – 150 de neuroni pe fiecare strat, activare RELU, stratul de ieșire – activare Softmax, funcția de pierdere Logarithmical categorical cross-entropy.

Pentru clasificarea nivelului de joc, am obținut următoarele valori pentru acuratețe:

- Scala cu 2 nivele – 98.72% utilizând un DNN cu 3 straturi ascunse – 300 de neuroni pe fiecare strat, activare RELU, stratul de ieșire cu 5 neuroni, activare Softmax, funcția de pierdere Logarithmical categorical cross-entropy.
- Scala cu 4 nivele – 98.67% utilizând un DNN cu 3 straturi ascunse – 150 de neuroni pe fiecare strat, activare RELU, stratul de ieșire cu 5 neuroni, activare Softmax, funcția de pierdere Logarithmical categorical cross-entropy.
- Scala cu 11 nivele – 98.75% utilizând un DNN cu 3 straturi ascunse – 150 de neuroni pe fiecare strat, activare RELU, stratul de ieșire cu 5 neuroni, activare Softmax, funcția de pierdere Logarithmical categorical cross-entropy.

Ulterior, am adus modificări la abordarea noastră și am adăugat clasificatori kNN, SVM, RF și LDA atât pentru determinarea nivelului de frică, cât și pentru determinarea următorului nivel de joc [10]. Am utilizat două metode pentru calculul acurateței clasificatorilor, dependent de utilizator/jucător și independent de utilizator/jucător.

Pentru clasificatorul de predicție a nivelului de frică, am obținut cea mai mare acuratețe de peste 98% cu validare încrucișată cu unul din algoritmi kNN sau RF, atât pentru modalitatea independent de jucător, cât și pentru modalitatea dependent de jucător. Aceeași tendință o găsim și în cazul clasificatorului pentru predicția nivelului următor de joc, unde cea mai mare acuratețe cu validare încrucișată a fost obținută cu RF, peste 99.75%.

Cele mai importante atribute pentru recunoașterea emoției au fost GSR, HR și valorile EEG în domeniul beta. Pentru predicția următorului nivel de joc, parametrul "nivelul de frică dorit" a influențat ieșirile clasificatorului.

Am evidențiat necesitatea dezvoltării de sisteme TREV cu recunoașterea emoțiilor utilizând o abordare centrată pe om în [6], [7]. Acest subiect este tratat în secțiunea "Abordarea

centrată pe om pentru sisteme TREV", unde sunt prezentate metodologii pentru construirea de astfel de sisteme având în centru terapeuții și pacienții.

Complexitatea sistemelor TERV poate fi abordată printr-o arhitectura a sistemelor bazată pe holoni [11]. Secțiunea "O Arhitectură bazată pe holoni pentru sisteme TREV" descrie propunerea noastră pentru sisteme pentru tratarea fobiilor, numită PhoVRET.

Modul în care ar trebui achiziționate biosemnalele pentru a obține date valide este descris în subcapitolul "Provocări privind integrarea recunoașterii automată a emoțiilor în sisteme TERV". Este definit conceptul de artefact ca fiind orice alterare în datele fiziologice ca urmare a unor acțiuni externe realizate de om, precum mișcări ale capului, mâinii, corpului, modificări fără nicio legătură cu efectele emoționale generate de anumiți stimului și am propus o metodă în trei pași pentru recunoașterea artefactelor [8]. Metoda constă în:

- Pasul I Măsurarea de referință a artefactelor.
- Pasul II Măsurarea artefactelor în timpul jocului.
- Pasul III Evaluarea potrivirii artefactelor.

Pentru validarea metodei propuse, am utilizat un al doilea joc de RV dezvoltat de echipa proiectului. La studiu au participat 5 subiecți, care au jucat jocul de mai multe ori și cărora le-am achiziționat semnalele fiziologice (GSR, HR și RR). Pentru HR și GSR am utilizat un dispozitiv Shimmer3 GSR+ și valorile RR au fost calculate în funcție de distanța dintre două dispozitive tracker HTC Vive.

Am verificat potrivirea între segmentele de referință cu artefacte și segmentele din timpul jocului prin calculul unor valori bias, eroarea medie absolută, eroarea procentuală absolută medie.

Rezultatele obținute au arătat că bias-ul este mai mic, dar nesemnificativ, pe segmentele de date aliniate față de segmentele înainte și după momentele cu artefacte. În cazul RR, valorile măsurate au fost identice în sesiunile de referință și de joc.

Având în vedere experiența acumulată în timpul experimentelor în laborator, am proiectat un protocol pentru achiziția semnalelor biofizice în medii de realitatea virtuală în scopul tratării fobiilor. Acest protocol a fost validat printr-un experiment la care au participat 7 subiecți. Am determinat cea mai influentă combinație de atribute extrase din EDA/HRV, dintr-un total de 32 de atribute. Valorile pentru SSE (Sum Squared Error – suma pătratelor erorilor) descresc după cum urmează: 3,827 (în cazul 1 atribut), 3,059 (2 atribute), 2,663 (3 atribute), 2,041 (4 atribute), 1,656 (5 atribute), 1,286 (6 atribute), and 1,031 (7 atribute).

Capitolul 4 cu titlul "Construirea de grupuri creative în învățarea colaborativă utilizând tehnici de inteligență artificială" se bazează pe articolele publicate de autoarea acestei teze de

abilitare în calitate de coautor în [20], [21], [22], [23]. Cercetarea a plecat de la o problemă reală întâlnită în procesul educațional și anume, cum putem defini echipe de studenți de la profilul informatică pentru a dezvolta cele mai creative soluții software la problemele din lumea reală. În prima parte a capitolului este prezentată importanța subiectului, accentuând necesitatea lucrului studenților în echipe. Întrebarea de cercetare care a ghidat studiul din capitolul 4 al tezei este: Poate fi accentuată creativitatea grupului de studenți prin asigurarea unui mediu instrucțional adecvat și prin organizarea optimă a studenților în grupuri utilizând tehnici ale inteligenței artificiale?

În secțiunea "Algoritm Q-Learning pentru construirea grupurilor creative de studenți", este prezentată o metodă bazată pe algoritmul Q-learning pentru a construi în mod optimal cele mai creative grupuri de învățare (GC-Q-Learning) introdusă în [20], [22], [23].

Algoritmul constă în:

1. Se construiește o matrice bi-dimensională Q pentru toate perechile posibile $\langle state, action \rangle$:

   $(c_1, c_2, \dots c_m, id\_group, action\_number, q)$

   O valoare pentru $action\_number$ egală cu *i* înseamnă că dacă un tip particular de student (descris prin vectorul său de creativitate $(c_1, c_2, \dots c_m)$) va fi mutat în grupul cu valoarea id_group egală cu *i*, atunci contribuția sa la creativitatea grupului este cuantificată prin q (în această etapă). Toate elementele din coloana q pot fi inițializate cu 0 sau cu o valoare random mică. Pe fiecare linie a matricei sunt incluse datele care corespund fiecărui tip de student, valorile pentru atributele studenților, numărul curent al grupului, numărul acțiunii, valoarea calculată pentru q *(care cuantifică un potential pentru creativitate)*. Un tip particular de student poate avea mai multe linii corespondente, câte una pentru fiecare combinație $\langle current\ id\_group, action \rangle$

2. Se inițializează *optimal_policy* cu o politică inițială. În cazul nostru, politica optimală este reprezentată de gruparea optimală a studenților care maximizează creativitatea grupului. Gruparea inițială este setată de instructor și studenți, experiența noastră arată că studenții tind să se grupeze pe baza afinităților inter-personale..

3. Se grupează studenții și se realizează o sesiune de lucru, în care este evaluată creativitatea fiecărui grup și o valoare este asignată recompensei $R(s, a)$. Valorile pentru $R(s, a)$ sunt obținute cu ajutorul experților umani. Recompensa descrie potențialul creativității grupurilor. Matricea Q este actualizată pentru fiecare sesiune de lucru cu procedura de mai jos.

procedure working_session_computation
select action of (optimal_policy) /* student grouping*/
compute $R(s,a)$
compute table Q

4. Se analizează creativitatea fiecărui grup având în vedere un obiectiv global (politica de grupare optima), care se apropie de valoarea maximă posibilă pentru R, pentru fiecare grup sau pentru toate grupurile. Se reia pasul 3 dacă este necesar.

Politica optimală este definită prin tupluri de forma $(c_1, c_2, \dots c_m, id\_group)$.
Următoarele notații au fost utilizate:

n – număr de studenți
$c = (c_1, c_2, \dots c_m)$ - vector de creativitate, c_i reprezintă un atribut al studenților ce influențează creativitatea grupului, m – număr de atribute individuale
id_group – identificator al grupului
k – numărul de grupuri
$(c_1, c_2, \dots c_m, id\_group)$ – o stare (*s*) compusă din vectorul de creativitate și identificarea grupului
acțiunea (*a*) – acțiunea de a muta un student în alt grup la a cărei creativitate ar putea contribui cel mai mult
Q – exprimă calitatea asocierii dintre stare și acțiune, în sensul scopului nostru, de a construi cele mai creative k grupuri
R - recompensa este valoarea creativității grupului și variază între 1 și 5.

Algoritmul a fost testat prin executarea mai multor cazuri de utilizare. Pentru studenți am considerat două atribute, nivelul individual de creativitate si nivelul de motivație. Pentru grupurile de studenți am calculat valorile Q. Rezultatele au arătat că algoritmul este o soluție validă pentru a grupa studenții asigurând o creativitate ridicată a grupurilor.
În subcapitolul "Un sistem multiagent pentru construirea grupurilor creative de studenți", descriem propunerea noastră pentru un sistem multiagent, numit GC-MAS și publicat în [20], [23]. Sistemul propus integrează algoritmul GC-Q-Learning. Cei cinci agenți care compun sistemul GC-MAS sunt: agentul de comunicare (CommGC); constructorul de grupuri creative (BuildGC); agentul de evaluare a creativității grupurilor (EvalGC); agentul stimulator (EnvrGC) și agentul facilitator (FcIGC).
Subcapitolul cu titlul "Clasificatori Bayes pentru construirea grupurilor creative de studenți" are ca scop prezentarea unui model și metode de a grupa studenții într-un mod optim în

situații de învățare colaborativă utilizând clasificatori Bayes [21]. Ideea principală a abordării noastre este considerarea caracteristicii grupului cea mai relevantă pentru scopul propus și repetarea grupării studenților pe baza valorilor unor atribute individuale. Modelul nostru conține 3 stagii prezentate în Figura 3.

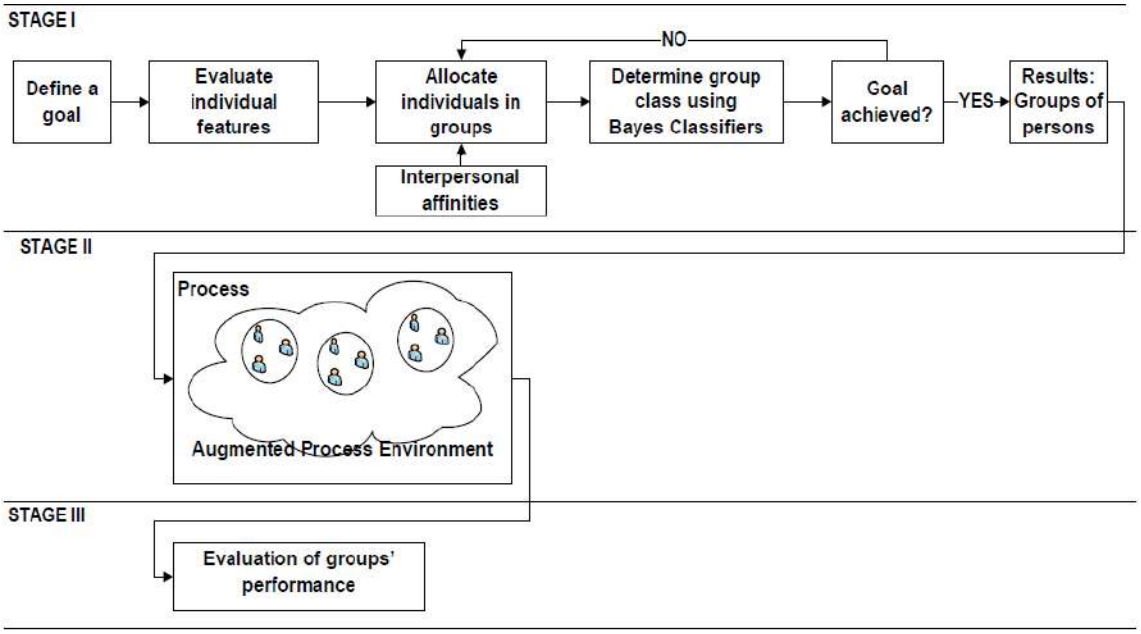

Figura 3. Model pentru construirea celor mai performante grupuri colaborative [21]

Modelul a fost validat într-un scenariu din lumea reală: scopul nostru a fost să grupăm 20 de studenți ai programului de studiu informatică în cele mai creative echipe. Am utilizat clasificatori Bayes, unul pentru predicția clasei de creativitate pentru fiecare student și altul pentru predicția clasei de creativitate pentru fiecare grup. Rezultatele obținute au confirmat o creștere a performanțelor la învățare a studenților prin construirea de echipe conform modelului nostru.

Capitolul 5, cu titlul "Recunoașterea automată a emoțiilor în procesul de învățare" prezintă rezultatele cercetării începută în anul 2023 în proiectul cu titlul Învățare Afectivă: Beneficii și Riscuri Etice în Învățământul Superior (Affective Learning: Benefits and Ethical Risks in Superior Education - ALBER) coordonat de autoarea acestei teze. Rezultatele cercetărilor au fost publicate în articolele [24], [25]. Capitolul începe cu o scurtă introducere privind importanța domeniului Învățare Afectivă. Am evidențiat rolurile emoțiilor în context academic și multitudinea lor: anxietate, speranță, lipsă de speranță, ușurare, plăcerea învățării, mândria succesului, furie, rușine, plictiseală, surpriză, tristețe, frustrare, confuzie, fericire, frică, bucurie, dezgust, interes, curiozitate, dispreț, încântare și entuziasm [26], [27], [28].

În subcapitolul "Roluri ale recunoașterii automată a emoțiilor în educație" sunt prezentate 6 categorii de roluri ale sistemelor RAE în educație [24]:

- Dezvoltarea Sistemelor Inteligente de Tutorat cu abilități emoționale, capable să detecteze emoțiile și să reacționeze adecvat la acestea.
- Sprijinirea angajării și motivării studenților și profesorilor.
- Evaluarea învățării (detectarea încercărilor de fraudă academică de către studenți).
- Evaluarea predării (profesorii trebuie să fie conștienți de impactul emoțiilor asupra procesului de predare-învățare).
- Construirea de medii de învățare adecvate procesului.
- Sprijinirea studenților cu nevoi speciale (suferind de ADHD, anxietate, etc.).

Pe baza setului de date DEAP, am construit un model 1D-CNN pentru recunoașterea a șapte emoții des întâlnite în context academic, plictiseală, confuzie, frustrare, curiozitate, entuziasm, concentrare și anxietate. Contribuția noastră a constat în obținerea unul model performant pe baza a numai a 5 canale EEG [25]. Modelul și rezultatele sunt descrise în subcapitolul "Model 1D-CNN pentru recunoașterea emoțiilor pe baza a 5 canale EEG – setul de date DEAP". Am plecat de la modelul RAE propus de Akter și alții (2022) care utilizează 14 canale EEG (FP1, FP2, AF3, Fz, F3, F4, F7, F8, FC1, C4, P3, P4, PO3, PO4) cu performanțele în ceea ce privește acuratețea: pentru valență - 99.89%, și pentru excitare – 99.83% [100].

Pentru a obține modelul RAE am folosit procedura din Figura 4.

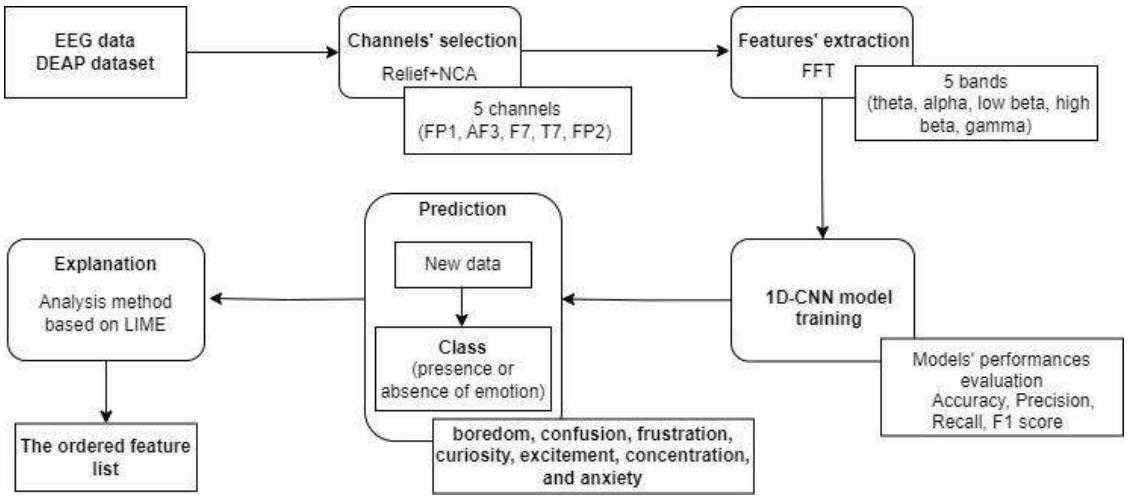

Figura 4. Procedura pentru obținerea modelului RAE în procesul de învățare [25]

Modelul nostru utilizează 5 canale din semnalele EEG și anume FP1, AF3, F7, T7, FP2. Am efectuat extragerea caracteristicilor utilizând FFT (Fast Fourier Transform), după care am aplicat standardizare.

Performanța modelului 1D-CNN bazat pe 5 canale EEG este prezentată în Tabelul 1. Timpul de antrenare a fost 9342.58 secunde și valorile pentru acurateță peste 99.21%. Scorul F1 variază de la 91.76%, in cazul anxietății, la 99.07%, în cazul concentrării.

Tabelul 1. Performanțele modelului 1D-CNN bazat pe 5 canale EEG [25]

Emotion	Performanță (%)				Timp de antrenare (secunde)
	Acuratețe test	Precizie	Rechemare	Scorul F1	
Plictiseală	99.64	94.62	97.41	95.97	8281.34
Confuzie	99.70	99.17	96.19	97.63	6004.65
Frustrare	99.66	98.98	95.13	96.97	5690.35
Curiozitate	99.80	99.74	96.83	98.24	8675.36
Entuziasm	99.91	98.74	97.75	98.24	5790.36
Concentrare	99.70	99.16	98.98	99.07	9342.58
Anxietate	99.21	95.03	88.96	91.76	6655.65

Am dezbătut subiectul eticii privind utilizarea RAE în procesul de învățare în subcapitolul cu titlul "Model etic pentru RAE în procesul de învățarea online". Modelul nostru etic, prezentat în [24] consideră 16 clase de riscuri etice asociate utilizării RAE în procesul de învățare:

- „Prejudecăți și discriminare,
- Rezultate nefiabile, incerte, nesigure sau slabe.
- Rezultate netransparente, inexplicabile, nejustificate sau total neprevizibile.
- Încălcarea vieții private prin (1) proprietatea și gestionarea incorectă a datelor cu caracter personal, (2) neacordarea și retragerea consimțământului și (3) supraveghere internă.
- Nedreptate și diviziune digitală.
- Înșelăciune.
- Manipularea și construirea de relații autoritare.
- Schimbări în percepția umană asupra realității, înțelegerii, expertizei și comportamentului natural.
- Portretizare eronată a ființelor umane și a emoțiilor.
- Negarea sau ocolirea autonomiei și drepturilor individuale (restricționarea capacității utilizatorilor de a-și exercita voința sau libertatea de exprimare, decizii nelibere și neinformate cu privire la utilizatori).
- Utilizare dual.
- Izolarea indivizilor, dezintegrarea conexiunilor sociale și dezumanizarea relațiilor interumane prin interacțiunea emoțională și socială cu sisteme de inteligență artificială de înaltă performanță, dar lipsite de conștientizare de sine.

- Dependență de o mașină.
- Riscul de a pierde simțul identității individuale.
- Înlocuirea profesorilor.
- Lipsa sustenabilității energetice."

Modelul etic (Figura 5) urmărește fluxul dedicat dezvoltării unui model de învățare automată cu considerarea în plus a eticii și a celor trei nivele definite în framework-ul lui Leslie din [29].

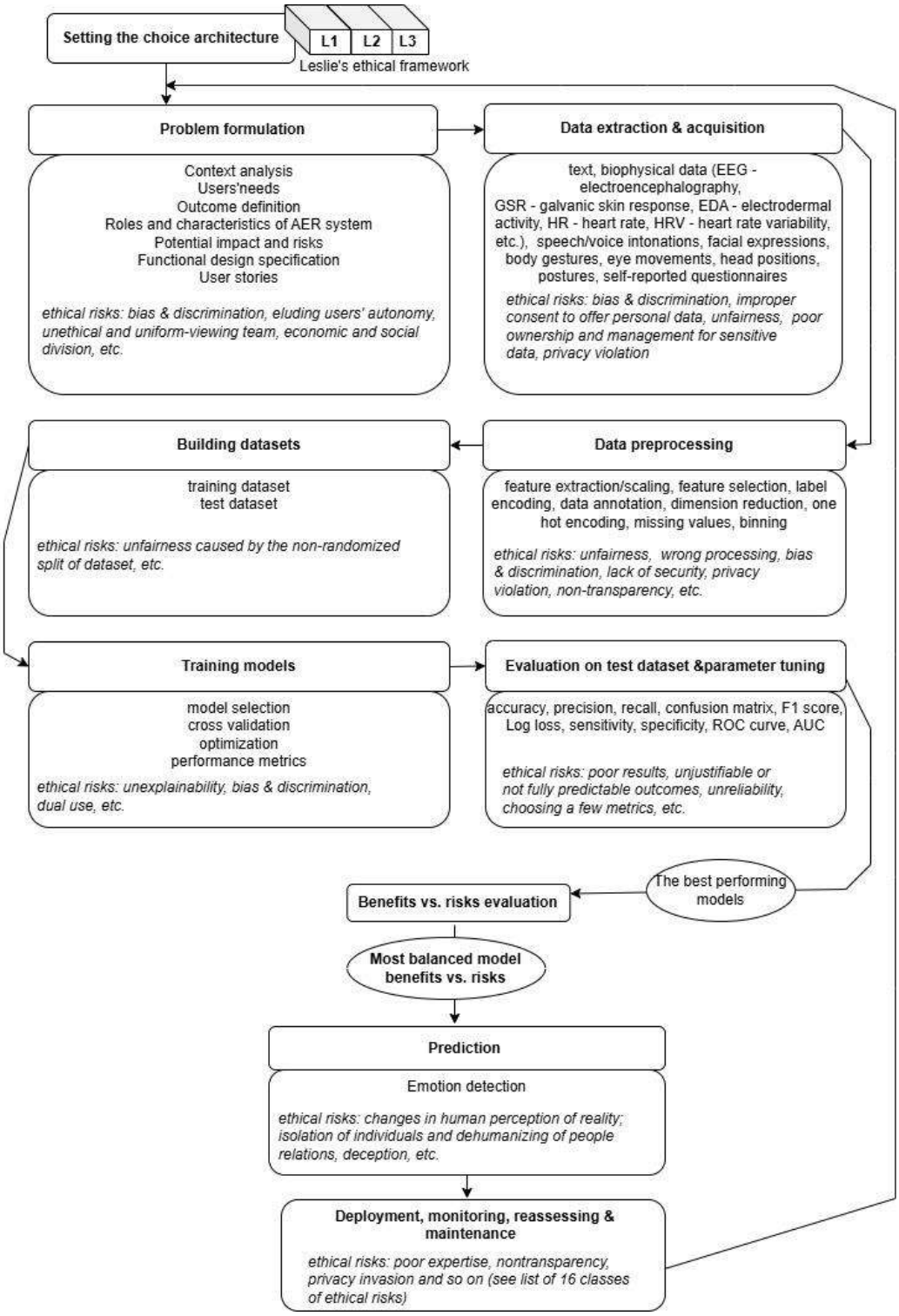

Figura 5. Model etic pentru RAE în învățarea online [24]

Capitolul 6 cu titlul "Explicabiliate în Învățarea Automată" este bazat pe articolele publicate de autoarea acestei teze de abilitare în calitate de coautor în [30] și [25]. În prima parte a capitolului este investigată importanța subiectului în peisajul aplicațiilor cu inteligență artificială. Accentuăm faptul că modelul conceptual FAT (fairness, accountability, and transparency – echitate, responsabilitate și transparență) din [31] trebuie operaționalizat atât în industrie, cât și în mediul academic. Ieșirile diferiților explicatori aplicați pentru aceeași instanță pot fi diferite. Mai mult, un explicator aplicat de mai multe ori unei instanțe poate genera rezultate diferite. Problema dezacordului în învățarea automată constă în obținerea de explicații contradictorii pentru același eșantion și model [32], [33], [34].

Propunem în [25] o metodă bazată pe LIME pentru a furniza explicații, care este descrisă în subcapitolul "O metodă de analiză cu LIME pentru explicații". Metoda cuprinde următorii pași:

1. "LIME este executat de mai multe ori (în experimentele noastre, de 20 ori).
2. La fiecare execuție se determină influența atributului asupra predicției (unele atribute sprijină prezența unei emoții specific, altele sprijină absența acelei emoții).
3. Pentru fiecare atribut, este calculat numărul de apariții ale atributului în subsetul de atribute care sprijină prezența unei emoții specifice, precum și numărul de apariții ale atributului în subsetul de atribute care sprijină absența unei emoții specifice.
4. Pentru fiecare atribut, se calculează valoarea absolută a diferenței dintre cele două frecvențe calculate anterior.
5. Ulterior, atributele sunt sortate în ordine descendentă pe baza diferenței absolute. Atributul cu cea mai mare valoare contribuie cel mai mult la prezența sau absența emoției. Dacă numărul de apariții ale unui atribut în subsetul de atribute care sprijină prezența emoției este mai mare decât numărul de apariții ale aceluiaș atribut în subsetul de atribute care sprijină absența emoției, atunci se consideră că acel atribut sprijină prezența emoției, altfel invers. [25]"

Metoda a fost validată pe predicțiile obținute cu modelul 1D-CNN cu 5 canale. Rezultatele arată funcționalitatea metodei în fiecare caz studiat.

De exemplu, într-un caz al predicției absenței plictiselii, sunt obținute următoarele rezultate pentru fiecare atribut rulând LIME de 20 de ori: culoarea orange marchează atributele care sprijină mai mult prezența emoției, culoarea albastră le marchează pe cele care sprijină absența emoției, iar atributele fără nicio culoare au diferența absolută egală cu 0.

19	6	23	7	4	11	12		2	5	13	20	24
14	12	12	10	8	8	8		6	6	6	6	6
1	9	14	18	21	0	8	15	16	17	22	3	10
4	4	4	4	4	2	2	2	2	2	2	0	0

Adunând diferențele absolute, putem concluziona că atributele influențează mai mult absența emoției (72) decât prezența acesteia (62).

Metoda este costisitoare din punct de vedere al timpului necesar, așa că intenționăm să oferim un nou algoritm bazat pe LIME.

În subcapitolul "O metodă de agregare bazată pe clustere pentru alinierea explicațiilor", este prezentată o metodă inspirată din raționamentul bazat pe cazuri pentru agrearea diferitelor explicații [30]. În elaborarea acestei metode a fost utilizată abordarea lui Pirie și alții (2023) pentru agrearea explicațiilor, care a aplicat o aliniere locală și a propus o metrică de încredere a alinierii între explicatori. De asemenea, aceștia au dezvoltat un framework pentru agrearea explicatorilor, numit AGREE (AGgregation for Robust Explanations) [35]. Metoda conține două etape: în prima etapă, sunt generate explicații și în etapa a doua este aplicată o metodă de agreare bazată pe clustere inspirată din Case-Based Reasoning.

Metoda a fost evaluată pe șase seturi de date des utilizate (Pima Indian Diabetes Dataset [36], Indian Liver Patient Dataset [37], Hepatitis Dataset [38], Fetal Dataset [39], Abalone Dataset [40], Water Quality Dataset [41]) prin furnizarea vectorilor cu ponderi ale explicațiilor agregate spațiului de caracteristici al unui clasificator k-NN ponderat și compararea performanțelor de predicție cu cele obținute cu un algoritm k-NN neponderat. Explicațiile au fost generate cu LIME, Anchors, Kernel SHAP, and Tree SHAP.

Rezultatele obținute validează strategia propusă (Tabelele 2 și 3).

Tabelul 2. Acuratețea algoritmului k-NN ponderat vs. acuratețea algoritmului k-NN neponderat [30]

	Diabetes	Liver	Hepatitis	Abalone	Water	Fetal
Average accuracy for weighted k-NN	**74.3**	**67.93**	**94.85**	**53.55**	**61.97**	**92.87**
Non-weighted k-NN	72.58	65.65	92.43	53.62	60.86	91.74

Tabelul 3. Scorul F1 pentru algoritmul k-NN ponderat vs. scorul F1 pentru algoritmul k-NN neponderat [30]

	Diabetes	Liver	Hepatitis	Abalone	Water	Fetal
Average F1 score for weighted k-NN	**60.67**	**78.08**	**71.51**	**46.44**	**72.35**	**83.04**
Non-weighted k-NN	55.74	76.45	55.01	47.78	42.81	79.95

Strategia este costisitoare din punct de vedere al resurselor și timpului de calcul. Intenționăm să ne dedicăm cercetării în XAI, dezvoltând metode fiabile cu costuri de calcul reduse.

Summary

A selection of the main scientific contributions of the author of this habilitation thesis are briefly presented in this abstract.

Chapter 1 entitled "A Look at Affective computing", introduces the research area of Affective Computing (AC), started by Prof. Picard (1995) [1]. In short, the field deals with the development of technologies that are aware of human emotions and has applications in healthcare, education, entertainment, virtual reality, recommendation systems, marketing, human-machines interfaces and so on. Affective Learning (AL) is a sub-domain of AC, which considers the inclusion of emotion-aware technologies in the learning process [2]. Affect-aware learning technologies (AALTs) represent those technologies which can recognize and show emotions in the learning situations. Mello&Graesser (2015) studied the AALTs and emphasize their benefits in the learning process [3].

We highlight the major concerns related to AC, namely, the ethics risks associated with the domain. The applications that contain emotion recognition using artificial intelligence are classified in the class of high-risk systems in AI Act (2024) and they have specific requirements [4].

AC comprises two major topics: emotion detection/recognition and emotion expression by machines. The research of the author of this thesis includes the recognition of emotions in the context of phobia treatment using virtual reality and in the context of learning.

Chapter 2 with title "Data for Automated Emotion Recognition" aims to provide a brief presentation of the emotion models and data used for emotion recognition. We present the discrete model, the dimensional model, and the componential model, of which the first two are used in this paper. For our research, we used three datasets containing biophysical data.

Two datasets represent our contribution and were used in our research published by the author of this habilitation thesis as co-author in [5, 6, 7, 8, 9, 10, 11, 12].

Dataset 1 contains EEG (Electroencephalography), EDA (Electrodermal activity) and HR (Heart Rate) signals captured from 4 subjects in both virtual and in real environment. We used Acticap Xpress Bundle device with 16 dry electrodes to capture signals from the channels FP1, FP2, FC5, FC1, FC2, FC6, T7, C3, C4, T8, P3, P1, P2, P4, O1 and O2. EDA and HR signals were captured with the GSR unit of a Shimmers Multi-Sensory device.

Dataset 2 contains GSR (Galvanic Skin Response), HR and RR (Respiration Rate) signals acquired from 5 subjects in two situations. In the first, we measured GSR, HR and RR as reference during the subjects performing the following actions deep breath, head movement

to the left, head movement to the right, head movement up, head movement down, click with the right hand on the HTC Vive controller, right hand raise. In the second, we performed the same measurements with the subjects while they were playing a VR-based game. For HR and GSR measurements, we used the Shimmer3 GSR+ Unit (https://www.shimmersensing.com/product/shimmer3-gsr-unit/) and the Respiration Rate was calculated according to the distance between two HTC Vive trackers.

The third dataset is a benchmark one, namely DEAP (Database for Emotion Analysis using Physiological signals) dataset [13].

Chapter 3 entitled "Emotion-Aware Virtual Reality Exposure Therapy Systems for Phobia Treatment" is based on the articles published by the author of this habilitation thesis as co-author [5], [6], [7], [8], [9], [10], [11], [12], [14], [15], [16]. In the first part of the chapter, we show the importance of the subject of Virtual Reality Exposure Therapy (VRET) in phobia treatment. Our studies focused on the acrophobia, which has a high incidence affecting 1 in 20 individuals according to [17]. Usage of VR technology in phobia treatment was first reported at the end of 1990 [18], since then many people preferring VRET over other treatment methods.

The sub-chapter "AER models based on DEAP datasets" presents our approaches to develop AER models based on the DEAP dataset.

In the first group of models, we recognized fear according to the 2 – level scale (absence/presence) and the 4-level scale (no/low/medium/high) paradigms [14]. We built five inputs sets for each paradigm: Raw, Power Spectral Density, Petrosian Fractal Dimension, Higuchi Fractal Dimension, Approximate Entropy for EEG and physiological recordings (hEOG, vEOG, zEMG, tEMG, GSR, Respiration, PPG and temperature). We tested more machine learning models: four deep neural networks and four ML models - Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Random Forest (RF) and k-Nearest Neighbors (kNN) and applied more feature selection - Principal Component Analysis (PCA), Sequential Feature Selector (SFS), Fisher selection. The most performant AER models were RF classifiers with no feature selection with accuracy – 93.13%, F1 score – 93.11% in the case of 2-level scale and accuracy – 85.74%, F1 score – 85.33% in the case of 4-level scale. The performances were achieved on the input set obtained from DEAP applying Power Spectral Density (PSD) of all 32 EEG channels in the alpha, beta and theta frequency ranges and computed the mean values for alpha, beta and theta PSD in the pre-frontal (FP), AF (between FP and F), frontal (F), FC (between F and C), central (C), temporal (T), P (parietal), CP (between C and P), O (occipital) and PO (between P and O) sides of the brain. Also, the input

set contained the 8 physiological features (hEOG, vEOG, zEMG, tEMG, GSR, Respiration, PPG and temperature) from DEAP. We extended in [16] the approach from [14] to recognize the 6 discrete emotions defined by Ekman in [19].

In [15], we considered for fear recognition only GSR and HRV signals from DEAP dataset and proposed a feature extraction protocol. The protocol consisted in segmenting each trial both in three non-overlapping windows and five overlapping windows. So, we extracted 33 types of features for EDA and 7 types of features for HRV, in total 40 type of features.

The pipeline for the process applied in [15] is shown in Figure 1.

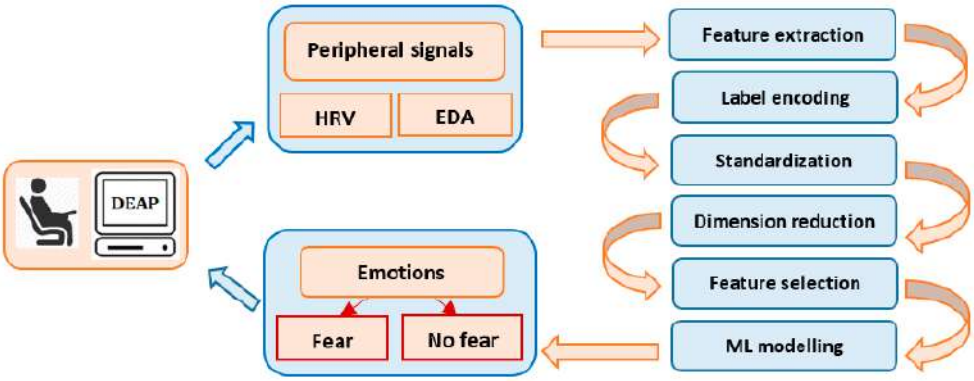

Figure 1. The pipeline of the process of automated fear recognition [15]

The non-overlapping approach consisted in dividing each segment of 60 seconds into three non-overlapping windows of 20 seconds each. In the overlapping approach, each segment was divided into five windows of 20 seconds each and 10 seconds overlapping other segments.

In both paradigms, overlapping and non-overlapping datasets, the best performances (over 89%) were achieved by SVM (Support Vector Machine) and GBT (Gradient Boosting Tree) algorithms. Considering ROC AUC score, we obtained the following best results:

- For the non-overlapping dataset: PCA reduction + SVM – 93.5%.
- For the overlapping dataset: GBT – 91.7%.

The sub-chapter "Development of the adaptive VRET systems with emotion recognition" presents our paradigm for the VR-based game for phobia treatment.

In the first VR-based game [6], [9], [14] the players were exposed in real-time to appropriate in-game height levels according to physiological data acquired from the players.

The architecture of the proposed VRET system is presented in Figure 2 [6], [9], [14]. We designed and trained two Deep Neural Networks (DNNs) to be integrated in the game. First DNN estimates the fear level of the players, and the second one sets the next game level to

be played by the user. In this way, the users were exposed gradually in real-time to the various height-levels according to their fear levels.

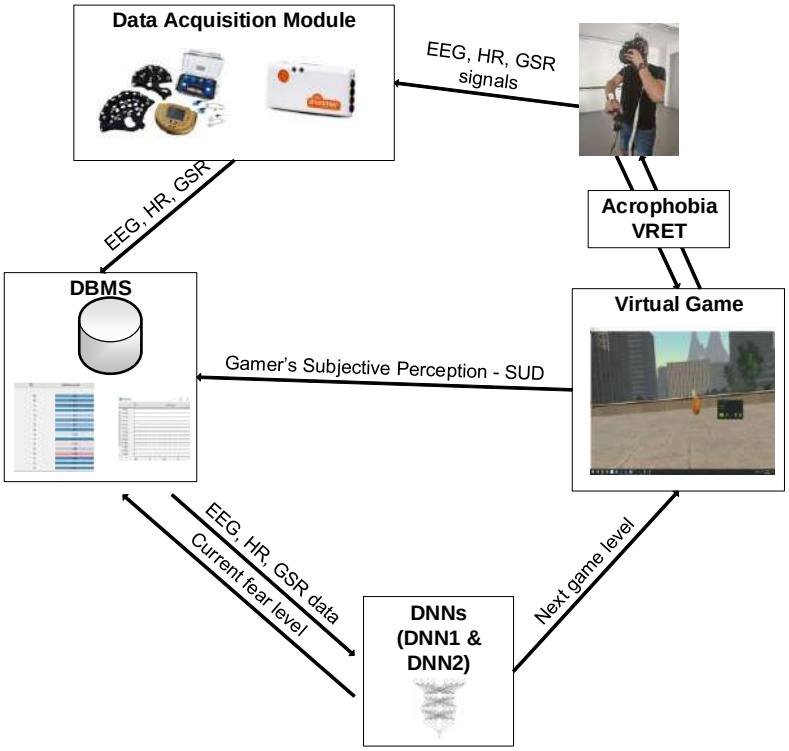

Figure 2. Architecture of the proposed emotion-aware VRET system [9]

The system comprises of a module for EEG, HR, and GSR signals acquisition. The signals are pre-processed and transferred to the DBMS (Database Management System). Data extracted from EEG, HR and GSR signals feed a deep neural network (DNN) classifier to determine the current level of fear. Using the same data and the desired fear level, the second deep neural network predicts the next game's level suitable for a specific user.

The work scenario for our game is presented in Figure 3. The user starts playing l_0 level, so the current level of game l_{cr} is l_0. The EEG, HR and GSR data are recorded during the game. When the user finishes a game level, the acquired data is fed to the deep neural network, DNN1, and the current fear level is predicted. The next desired fear level is computed according to the 2-choices scale or 4-choices scale paradigms.

In the case of 2-choices scale, we determined the desired fear level according to:

if $fl_{cr} == 0$ then $fl_d = 0$

if $fl_{cr} == 1$ then $fl_d = 1$

In the case of 4-choices scale, we determined the desired fear level according to:

if $fl_{cr} == 0$ or $fl_{cr} == 1$ then $fl_d = fl_{cr} + 1$

if $fl_{cr} == 2$ then $fl_d = fl_{cr}$

if $fl_{cr} == 3$ then $fl_d = fl_{cr} - 1$

The desired fear level and biophysical data are inputs for the second deep neural network (DNN2) and a game level is predicted (l_{pr}). The users play the predicted level of the game and their biophysical data are again recorded. The algorithm stops when a predefined number of epochs is reached.

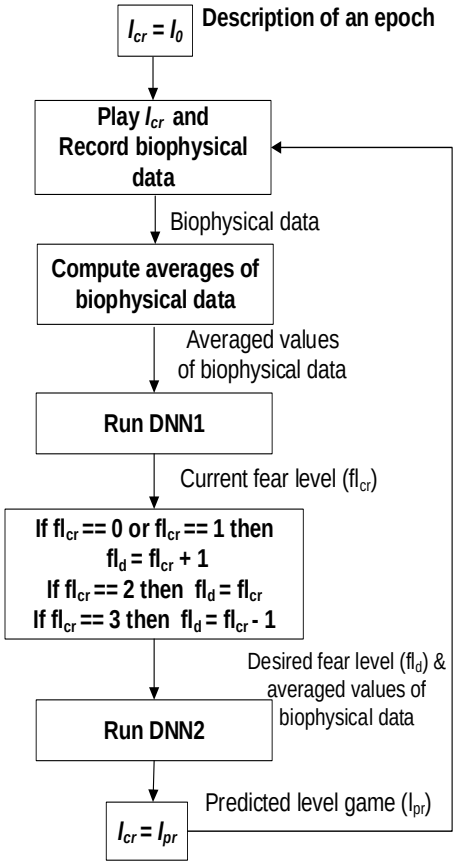

Figure 3. The algorithm in the case of 4-choices scale [9]

We tested three types of choices scale for fear assessment (2-choices scale, 4- choices scale and 11-choices scale) and achieved the following high accuracy for fear classification:

- 2 – choices scale 95.51% using a DNN with 3 hidden layers – 300 neurons on each layer, RELU activation, Output layer – Sigmoid activation, Binary cross-entropy loss function.
- 4 – choices scale 90.49% using a DNN with 3 hidden layers – 300 neurons on each layer, RELU activation, Output layer -Softmax activation function, Logarithmical categorical cross-entropy loss function.
- 11 – choices scale 85.09% using a DNN with 3 hidden layers – 150 neurons on each layer, RELU activation, Output layer -Softmax activation function Logarithmical categorical cross-entropy loss function.

For the game level classification, we achieved the following accuracy:

- 2 – choices scale 98.72% using a DNN with 3 hidden layers – 300 neurons on each layer, RELU activation, Output layer with 5 neurons, Softmax activation function, Logarithmical categorical cross-entropy loss function.
- 4 – choices scale 98.67% using a DNN with 3 hidden layers – 150 neurons on each layer, RELU activation, Output layer with 5 neurons, Softmax activation function, Logarithmical categorical cross-entropy loss function.
- 11 – choices scale 98.75% using a DNN with 3 hidden layers – 150 neurons on each layer, RELU activation, Output layer with 5 neurons, Softmax activation function, Logarithmical categorical cross-entropy loss function.

Subsequently, we modified our approach and added kNN, SVM, RF, and LDA classifiers for both fear level and the game level that should be played next [10]. We used two modalities to calculate the accuracies of the classifiers: a user-dependent and a user-independent.

With respect to the classifier predicting fear level, we achieved the highest cross-validation accuracy (over 98%) using either the kNN or RF algorithms, for both the player-independent and player-dependent modalities. The same trend occurred for the classifier predicting the game level, where very high cross-validation accuracies were recorded by the RF classifier (over 99.75%).

The most important features for AER were GSR, HR and the values of the EEG in the beta range. For next game level prediction, the parameter "target fear level" influenced the outputs of the classifier.

We stressed the necessity of using a human-centred approach for VRET systems development [6], [7]. We covered the subject in section "The Human-centred approach for VRET systems" and presented methodologies to build such systems having in the core therapists and patients.

The complexity of VRET systems can be managed with a holonic-based architecture [11]. The section "A holonic-based architecture for VRET systems" describes our holonic-based approach, called PhoVRET.

The way in which the biosignals should be acquired to obtain valid data is considered in sub-chapter "Challenges for integrating AER in VRET systems". We defined "artefacts as any misleading or confusing alterations in physiological data that appear as a result of external action such as head, hand or body movements, being unrelated to the emotional effects that specific stimuli or the object under observation exert upon the user" in [8] and proposed a method in three steps to artefacts' recognition. The method consists in:

- Step I Reference artefact measurement.

- Step II Artefact measurement during gameplay.
- Step III Artefact matching evaluation.

To validate our method, we used a second VR-based game developed in house. 5 participants to our study played several times the game and we captured physiological signals (GSR, HR and RR). Heart Rate and Galvanic Skin Response are measured by the Shimmer3 GSR+ and the Respiration Rate was computed according to the distance between two HTC Vive trackers.

We verified the matching between reference artefacts segments and data recorded in the game segments by computing the bias or Mean Absolute Error (MAE) or the Mean Absolute Percentage Error (MAPE) for GSR and HR. The results obtained from the laboratory experiments showed the bias is lower, but not significantly, on the aligned data segments than on the segments before and after the moments with artefacts. The deep breath artefact is more influencing than the other artefacts. Related to RR, the measured values were identical in both the reference and gameplay session.

We designed a protocol to record the biophysical signals in VR environments for phobia treatment according to our experience gained during our experiments [12]. To validate the protocol, we conducted an experiment involving 7 subjects. We found out the most influent combinations of features extracted from EDA/HRV (in total 32 features) based on regression analysis. The Sum Squared Error Estimation decreased as follows: 3.827 (in the case of 1 feature), 3.059 (2 features), 2.663 (3 features), 2.041 (4 features), 1.656 (5 features), 1.286 (6 features), and 1.031 (7 features).

Chapter 4 entitled "Building Creative Groups in Collaborative Learning Using Artificial Intelligence Techniques" is based on the articles published by the author of this habilitation thesis as co-author [20], [21], [22], [23]. The research was raised starting from a real problem, encountered in the educational process, namely, how we build teams of CS students capable of providing creative software solutions to the real-world problems. The importance of the topic is presented in the first part of the chapter, highlighting the necessity of involving students in teamwork-based tasks to be able to face the complexity of the real-world situations. Our research questions which guided the study from chapter 4 is: Can student group's creativity be enhanced by providing contextual instructional environments and organizing individuals into appropriate groups using artificial intelligence techniques?

In section "Q-Learning algorithm for building creative students' groups", we present the Q-learning algorithm - based method to build in an optimal way the most creative learning groups (GC-Q-Learning) introduced in [20], [22], [23].

The algorithm consists in:

1. Build a bi-dimensional matrix Q for all possible pairs $\langle state, action \rangle$: $(c_1, c_2, \dots c_m, id\_group, action\_number, q)$

   A value of the $action\_number$ equal to *i* means that if a particular type of student (given by their creativity vector $(c_1, c_2, \dots c_m)$) will be moved to the group having the value of id_group equal to *i*, then their contribution to group creativity is quantified by q (in this stage). All the elements in the q column may be initialized with 0 or with a randomly chosen low value. On each line of the matrix, the data that corresponds to each type of student involved in the grouping process is included, i.e. the values of their characteristics, the current group number, the action number, and the value computed for q *(that quantifies a potential for creativity)*. One particular type of student could have more corresponding lines, one for each combination $\langle current\ id\_group, action \rangle$
2. Initialize the *optimal_policy* with an initial policy. In our case, the optimal policy is an optimal grouping of students that maximizes group creativity. The initial grouping is set by the instructor and the students together and experience shows that they tend to group as cliques based on their inter-personal affinities.
3. Group the students and have them carry out working sessions, in which each group's creativity is assessed, and its score is assigned to the reward $R(s, a)$ The values of $R(s, a)$ are obtained with help from human experts. The reward describes the potential of the groups' creativity. Then, the matrix Q is updated for each such working session. This procedure is presented below.

   procedure working_session_computation
   select action of (optimal_policy) /* student grouping*/
   compute $R(s, a)$
   compute table Q
4. Analyse the group creativity for each group against the global objective (the optimal grouping policy), which is getting closer to the maximum value possible for R, for each group or for all the groups. Re-iterate from step 3, if necessary.

The optimal policy is defined by tuples $(c_1, c_2, \dots c_m, id\_group)$.

The following notations were used:

n – number of students

$c = (c_1, c_2, \dots c_m)$ - creativity vector, c_i represents a characteristic of students influencing group creativity, m – number of individual characteristics

id_group – the group identification

k – the number of groups

$(c_1, c_2, \dots c_m, id\_group)$ – a state (*s*), composed by the creativity vector and the group identification for each student

action (*a*) – the action of moving a student to another group in which he would contribute the most to increasing the group creativity

Q – expresses the quality of association between a state and an action, in the sense of our goal, to build the most creative k groups

R - reward is the value of group's creativity, and it ranges between 1 and 5.

We tested the algorithm performing more uses cases. We considered two features for students, creativity and motivation level, and computed the Q-values for groups of students. The results show that the algorithm provides a valid solution to group students in the best way with regards to group creativity

In the sub-chapter "A multiagent system for building creative students' groups", we describe our proposal for a multiagent system, namely GC-MAS [20], [23]. The system integrates the GC-Q-Learning. The five agents composing GC-MAS are: the Communication Agent (CommGC); the Creative Groups' Builder (BuildGC); the Creativity Evaluation Agent (EvalGC); the Creativity Booster (EnvrGC); and the Facilitator Agent (FclGC).

The sub-chapter entitled "Bayes classifiers for building creative students' groups" aims to present a model and a method to group students in the best teams in collaborative learning situations using Bayes classifications [21]. The main idea of our approach is to consider a group characteristic that is the most relevant for the proposed goal and to maximize it by repeatedly grouping people based on the values of some individual features. Our model comprises 3 stages which are presented in Figure 3.

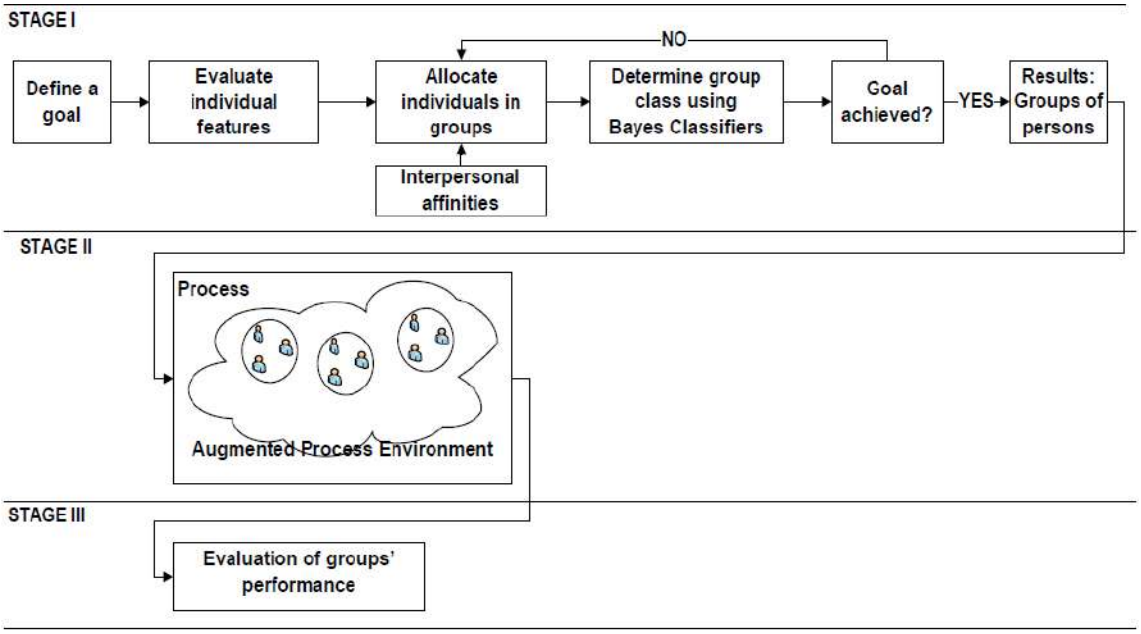

Figure 3. Model for building –the best collaborative groups [21]

We validated our model in a real-word scenario: our goal was to group 20 students of CS study program in the best creative teams. We used two Bayesian classifiers: one to predict creativity class for each student and other to predict the creativity class for each group. The obtained results confirm an increase of the learning performance of students grouped according to our model.

The **Chapter 5**, entitled "Automated Emotion Recognition in the Learning Process" presents the results of the research started in 2023 in the project entitled Affective Learning: Benefits and Ethical Risks in Superior Education (ALBER) conducted by the author of this thesis. The findings related to usage of AERs in educational systems are published in [24], [25]. The chapter starts with a short introduction of the importance of the Affective Learning domain. We highlight the emotions' roles in the academic context and the range of them: anxiety, hopefulness, hopelessness, relief, enjoyment of learning, pride of success, anger, shame, boredom, surprise, sadness, frustration, confusion, happiness, fear, joy, disgust, interest, curiosity, contempt, delight and excitement [26], [27], [28].

In the sub-chapter "Roles of AER in education", we present six categories of roles of AER systems in education identified in [24]:

- Development of Intelligent Tutoring Systems (ITS) with emotional abilities, able to detect emotions and react appropriately (the relationships between students' emotions, motivation, cognition, learning styles are speculated).
- Supporting engagement and motivation of learners and teachers (both students and teachers need to be aware of their emotional and mental states).

- Learning assessment (detect cheating by students).
- Teaching assessment (teachers need to be aware of the impact of emotions on the teaching-learning process).
- Building comfortable learning environments (increasing the efficiency of learning and teaching).
- Supporting students with special needs (ADHD, anxiety and so on) .

On top of the DEAP dataset, we built a 1D-CNN model for the seven emotions recognition, boredom, confusion, frustration, curiosity, excitement, concentration, and anxiety. Our contribution was obtaining a highly accurate model using only 5 EEG channels from DEAP dataset [25]. The model and the results are described in sub-chapter entitled "1D-CNN model for emotion recognition based on 5 EEG channels - DEAP dataset". We started from the AER model proposed by Akter et al. (2022) based on 14 EEG channels (FP1, FP2, AF3, Fz, F3, F4, F7, F8, FC1, C4, P3, P4, PO3, PO4) with the accuracies: for valence - 99.89%, and for arousal – 99.83% [100].

We used the framework from Figure 4 to obtain the AER model for the learning process.

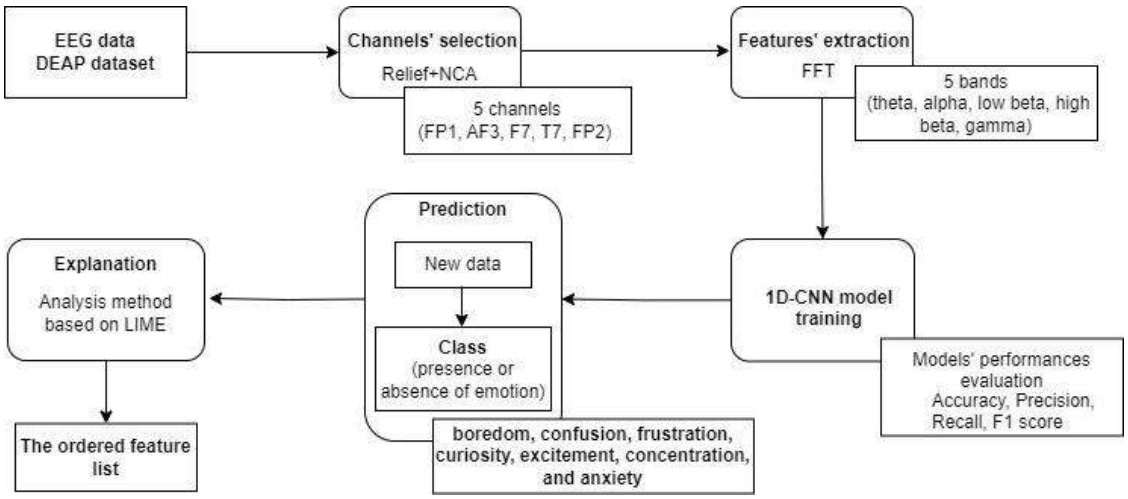

Figure 4. Our framework for AER in the learning process [25]

Our AER model uses five channels from EEG signals, namely FP1, AF3, F7, T7, FP2. We performed features extraction based on FFT (Fast Fourier Transform), the features' values were standardized by removing the mean and scaling to unit variance.

The performance of the 1D-CNN model based on 5 EEG channels, namely FP1, AF3, F7, T7, FP2 (Table 1). The running times are under 9342.58 seconds for all emotions' cases, and there can be noticed that all the test accuracy values are over 99.21%, The F1 score varies from 91.76%, in the case of anxiety to 99.07%, in the case of concentration.

Table 1. The performance of the 1D-CNN model based on 5 EEG channels [25]

Emotion	Performance (%)				Running time (seconds)
	Test accuracy	Precision	Recall	F1_score	
Boredom	99.64	94.62	97.41	95.97	8281.34
Confusion	99.70	99.17	96.19	97.63	6004.65
Frustration	99.66	98.98	95.13	96.97	5690.35
Curiosity	99.80	99.74	96.83	98.24	8675.36
Excitement	99.91	98.74	97.75	98.24	5790.36
Concentration	99.70	99.16	98.98	99.07	9342.58
Anxiety	99.21	95.03	88.96	91.76	6655.65

We covered the subject of ethics in usage of AERs in the learning process in the sub-chapter with title "Ethical model for AER in online learning". Our ethical model introduced in [24] considers 16 classes of ethical risks associated with usage of AER in the learning process:

- "Bias and discrimination,
- Unreliable, unsure, unsafe or poor results.
- Non-transparent, unexplainable, unjustifiable or not fully predictable outcomes.
- Privacy invasion by (1) inaccurate ownership and management of personal data, (2) failure in giving and withdrawing consent and (3) domestic surveillance.
- Unfairness and digital division.
- Deception.
- Manipulation and building authoritarian relations.
- Changes in human perception of reality, understanding, expertise and natural behaviour.
- Erroneous portraying of human beings and emotions.
- Denial or bypassing of individual autonomy and rights (restriction on users' ability to exercise free will or free speech, non-free and non-informed decisions regarding users, denial of right against self-incrimination).
- Dual use.
- Isolation of individuals, disintegration of social connections and dehumanizing of people relations by emotional and social interaction with high-performance, yet lacking self-awareness, AI systems.
- Dependence on a machine.
- Risk of losing the sense of individual identity.
- Replacement of the teachers.

- Lack of energetic sustainability".

Our model (Figure 5) follows the ML model creation pipeline with the additional consideration of ethics and the three levels defined in the Leslie's ethical framework for AI from [29].

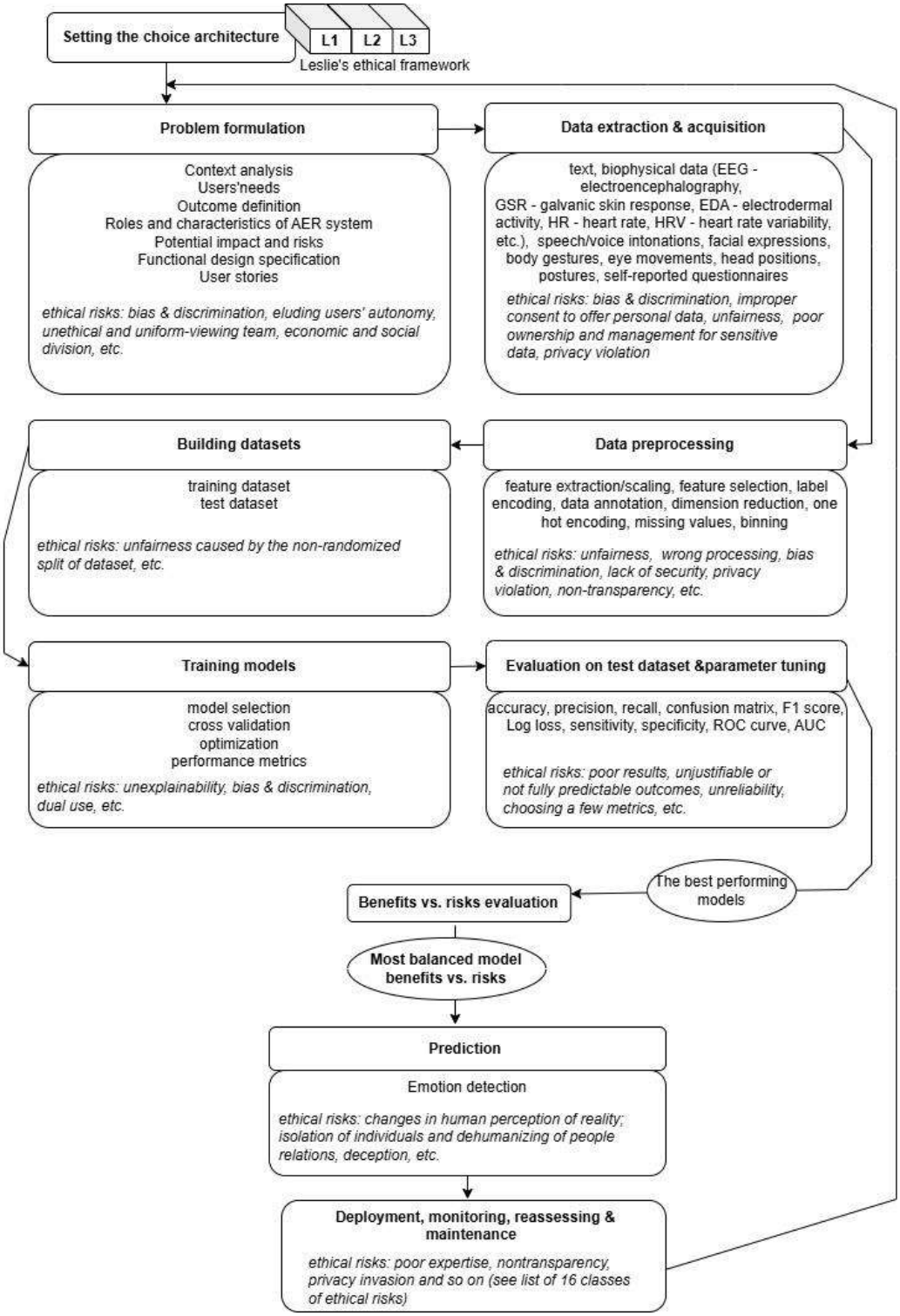

Figure 5. Ethical model for AER in online learning [24]

Chapter 6 with title "Explainability in Machine Learning" is based on the articles published by the author of this habilitation thesis as co-author in [30] and [25]. In the first part of the chapter, we make an investigation related to the importance of the subject in the AI-based

applications landscape. We stress that the FAT (fairness, accountability, and transparency) conceptual model from [31] is required to be considered and operationalized both in industry and academia. Also, we emphasize the issue related to the reliability of the explanation methods. The outputs of various explainers applied at the same instance are different. Furthermore, the same explainer running on the same instance generates different results. The disagreement problem in ML consists in obtaining contradictory explanations for the same sample and model [32], [33], [34].

We proposed in [25] a LIME-based method to provide explanations and it described in the sub-chapter entitled "A LIME-based analyse method for explanations". The method comprises the following steps:

6. "LIME is run more times (in our experiments we have run it 20 times).
7. For each run, the influence of the features on the prediction is obtained (some features support the presence of a specific emotion, other features support the absence of that emotion).
8. For each feature, the number of occurrences of the feature in the subset of features, which support the presence of a specific emotion, is computed, as well as the number of occurrences of the feature in the subset of features, which support the absence of the that emotion.
9. For each feature, the absolute difference of the two frequencies from above is calculated.
10. Afterwards, the features are sorted in a descending order, based on their absolute differences; the feature with the highest absolute difference is the most contributing to predict the presence or absence of an emotion. If the number of appearances of the feature in the subset of features which supports the presence of the emotion is higher than the number of appearances of the same feature in the subset of features which supports the absence of the emotion, then we consider that the feature supports the presence of the emotion, otherwise vice versa. [25]"

We validated our method on the predictions for emotion recognition obtained with the 1D-CNN model on 5 channels. The results show the functionality of the method in each studied case.

For example, in one case of boredom absence prediction, we obtained the following results for each feature running 20 times Lime algorithm. With the orange colour are marked the features which contribute several times to the presence of an emotion and with blue the

features which contribute several times to the absence of an emotion. The features with no colours have the absolute difference 0.

19	6	23	7	4	11	12		2	5	13	20	24
14	12	12	10	8	8	8		6	6	6	6	6
1	9	14	18	21	0	8	15	16	17	22	3	10
4	4	4	4	4	2	2	2	2	2	2	0	0

Summing the absolute differences, we can conclude that the features influence more the absence of boredom (72) than its presence (62).

The method is still expensive in the terms of time consumed, so we intend to provide a new LIME-based algorithm.

In the sub-chapter with the title "A cluster-based aggregation method for aligning explanations", we present a method inspired by Case-Based Reasoning to aggregate different explanations [30]. We used the Pirie et al. (2023) approach for explanations agreement, which applied the local alignment and proposed a Case Alignment Confidence metric between explainers and developed a framework for explainers' aggregation, named AGREE (AGgregation for Robust Explanations) [35]. The method comprises two stages: in the first stage, there are generated the explanations and in the second stage we applied a cluster-based aggregation method inspired from Case-Based Reasoning.

We assessed the method on six popular datasets (Pima Indian Diabetes Dataset [36], Indian Liver Patient Dataset [37], Hepatitis Dataset [38], Fetal Dataset [39], Abalone Dataset [40], Water Quality Dataset [41]) by providing the aggregated explanation weight vectors to the feature space of a weighted k-NN classifier (for each alignment scheme) and comparing the prediction performance against a non-weights k-NN algorithm, having as task a binary classification. We considered the explanations generated with LIME, Anchors, Kernel SHAP, and Tree SHAP.

The results we obtained confirm our strategy (Table 2, 3).

Table 2. Accuracy scores of the weighted k-NN algorithm comparing with the non-weighted k-NN classifier [30]

	Diabetes	Liver	Hepatitis	Abalone	Water	Fetal
Average accuracy for weighted k-NN	**74.3**	**67.93**	**94.85**	**53.55**	**61.97**	**92.87**
Non-weighted k-NN	72.58	65.65	92.43	53.62	60.86	91.74

Table 3. F1 scores of the weighted k-NN algorithm comparing with the non-weighted k-NN classifier [30]

	Diabetes	Liver	Hepatitis	Abalone	Water	Fetal
Average F1 score for weighted k-NN	**60.67**	**78.08**	**71.51**	**46.44**	**72.35**	**83.04**
Non-weighted k-NN	55.74	76.45	55.01	47.78	42.81	79.95

Still, the strategy is expensive in terms of computational resources and time. We intend to dedicate to the research in XAI developing reliable methods with low computational cost.

(B)Scientific and professional achievements and the evolution and development plans for career development

(B-i) Scientific and professional achievements

Introduction

I graduated in 1992 with an MsC in Computer Science from the Faculty of Mathematics, University of București. From 1992 to 2000 I acted as a software developer and head of the Computer Science department in IT sectors and industry. I have started my academic career in 2000 as an assistant professor at the Petroleum-Gas University of Ploiești, Informatics Department. In 2008 I got my PhD in Automatic Control with the thesis entitled Contributions to the Modelling and Controlling the Online Instructional Process Using Artificial Intelligence Techniques, supervisor Prof. Stelian Dumitrescu.

Currently, I am an associate professor and member of the department council at the Department of Computer Science, Information Technology, Mathematics, and Physics (iTIMF), Petroleum-Gas University of Ploiești.

From 2009 to 2024 I was involved in academic management, being the head of Computer Science Department (2009-2011), Department of Information Technology, Mathematics, and Physics (2011-2012) and Department of Computer Science, Information Technology, Mathematics, and Physics (2012-2024).

In the period 2016-2024, I acted as a member of the university senate with a short period (2016-2017) as chair of the educational committee. I am a member of the Coordinating Council CerTIMF Research Center (since 2013) and coordinator of the Computer Science&Education Laboratory (since 2019), Petroleum-Gas University of Ploiești. From 2016, I am evaluator of The Romanian Agency of Quality Assurance in Higher Education, Informatics field. I am also responsible for the Computer Science undergraduate program in our university since 2012.

As a teacher I have given lectures and/or laboratory classes for the subjects: Machine Learning, Computer Networks, Graphs Algorithms, Cryptography and Information Security, Data Analysis, eLearning in the Information society, Research Activities in Computer Science, Databases.

I have participated in the organization of many scientific events in the Petroleum-Gas University of Ploiesti as a member or coordinator of the organizing team for:

- The Computer Science and Education Symposium 2017, 2019, 2022, 2024.
- The Workshop Affective Learning: Benefits and Ethical Risks in Superior Education – ALBER – 2024 carried out by Research Grant GO-GICS, 11061/08.06.2023.
- The 14th International Workshop on Differential Geometry and Its Applications 2019, July 9th-1th, 2019.
- Math Pole Workshop carried out by CNFIS-FDI-2020-0034 project entitled Support for research excellence in STEM disciplines, 2020.
- Scientific Workshop Trends in the Computer Science Research for students in the master program 2022, 2023, 2024.

In 2024, I was member in organizing committee and chair in the program committee of the Workshop Multimodal, Affective and Interactive eXplainable AI, 19th – 20th October 2024, part of the 27th European Conference on Artificial Intelligence (A rank Core Conference), Santiago de Compostela, Spain.

Also, I coordinated the organization of many students' events and contests, and I participated in many events to promote the university's study programs.

As far as the IT industry is concerned, I am responsible on behalf of the Petroleum-Gas University of Ploiesti for the partnership contracts between the university and 7 companies.

I coordinated one research project in the period 2023-2024 with the title

- Affective Learning: Benefits and Ethical Risks in Superior Education (ALBER) – GO-GICS no. 11061/08.06.2023, 40000 lei, Petroleum-Gas University competition (http://csed.upg-ploiesti.ro/alber/)

and I was member in three research projects:

- Research, design, development and bench testing of a hybrid reactive propulsion system for space launcher applications area, with innovative inhomogeneous fuel and adaptive electrical control of parameters, ROSA-462-2017.
- Intelligent, wireless system of sensors for heat treatment equipment monitoring, PN-III-P2-2.1-CI-2017-0287.
- Intelligent system for monitoring the wearing of high voltage electrical protection equipment using RFID and wireless technologies, PN-III-P2-2.1-CI-2017-0262.

I was manager of two institutional development projects in 2019 and 2020 contributing to the development of the university's infrastructure through the development of teaching and research spaces and the purchase of equipment totalling over 500000 lei:

- 2020 CNFIS-FDI-2020-0034 - POLE4R&D – Support for Research in STEM, Research Support axis.
- 2019 CNFIS-FDI-2019-0066 - UPG-HUB4.0 – Multidisciplinary Research, Development and Innovation Hub, Research Support axis.

Currently, I am member of the project entitled Strategies for Creating Equitable Workplaces in Society 5.0, in which I have to propose a ML-based solution to extract information from several documents located in open databases and to formulate answers to different questions related to GDPR compliance in the labour relations.

After obtaining my Phd (2008), I published 58 research papers as co-author in scientific journals and conference proceedings. Among this, 5 in A journals, 1 in B journal, 3 in C journals and 7 in A conference's proceedings, 1 in a workshop associated with an A* conference, 6 in C CORE conferences or LNCS.

The A journals are:

- Artificial Intelligence Review
- Sensors

The B journal is Symmetry.

The C journals are:

- International Journal of Computational Intelligence Systems
- International Journal of Computers, Communications & Control

The A conferences are:

- the 28th International Conference on Information Systems Development (ISD 2019 France)
- the 27th European Conference on Information Systems (ECIS 2019 Sweden)
- the 27th International Conference on Information Systems Development (ISD2018 Sweden)
- International Conference on Cooperative Information Systems (CoopIS) 2017 (Rhodes, Greece)

The workshops associated with A* conferences are:

- International Workshop on Software Engineering in Healthcare Systems, ICSE '18: 40th International Conference on Software Engineering, Gothenburg Sweden
- HCML Perspectives Workshop at CHI Conference on Human Factors in Computing Systems, 2019 – paper without DOI.

I published 1 chapter in an A book, Academic Press, 1 chapter in a B book, Springer, and 1 chapter in a C book, InTech (according to SENSE). Also, I published as author or co-author 10 books (Romanian).

I served in the Program Committee for several conferences including ACM ITICSE conference (A conference, 2018-2025) and ACM SIGCSE (A conference, 2017-2022), and the MAI-XAI 24 Workshop, the European Conference on Artificial Intelligence (2024). Also, I was reviewer for more journals and conferences including IEEE Access, Information Fusion, IEEE Journal of Biomedical and Health Informatics, Scientific Reports, Cognitive Systems Research. I was an Editor of the journal Bulletin of Petroleum-Gas University, Series: Mathematics, Computer Science and Physics.

I am a member of the coordinating committee of a PhD student, National University of Science and Technology POLITEHNICA of Bucharest.

My research papers are cited in highly ranked journals and books:

A* forum – 6 citations

A forum – 48 citations

B forum – 76 citations

C forum – 59 citations

D forum – 115 citations

Books – 10 citations

PhD thesis, research reports, and other publications - 83 citations

My ISI Web of Knowledge Hirsch index is 6, the Scopus Hirsch index is 7, and the Google Scholar Hirsch index is 12.

My main research directions are related to: Affective computing (AC), Computer Science & Education, Ethics and Explainability in Machine Learning (ML). As I have taught courses in different areas, I have also tackled other research areas such as Databases, Encryption and Information Security.

Currently I am the leader of a small team (3 members) studying explainability and the disagreement problem in machine learning.

In this habilitation thesis, there are presented the main research contributions to the areas mentioned above and the plans for professional career development. So, the research presented in this thesis mainly focuses on the Automated Emotion Recognition (AER) with emphasis on the fear [5], [6], [7], [8], [9], [10], [11], [12], [14], [15], [16], to develop virtual reality (VR)-based applications for phobia treatment and emotions often encountered in the

learning process, boredom, confusion, frustration, curiosity, excitement, concentration, anxiety [25].

As a continuation of the research initiated in the PhD thesis, building the most creative groups of learners using ML techniques is a topic that interested the author of this thesis [20], [21], [22], [23].

We revealed the ethical concerns and we emphasized more issues on usage AER in learning process [24]. To develop ethical emotion-aware technology [24] we studied the explainability in ML and proposed an algorithm to "agree" different explanations generated with the same interpretability method [25]. The disagreement problem in machine learning is an important topic. To "agree" explanations obtained with different methods, a cluster-based aggregation method to align explanations provided with various interpretability methods is presented in [30].

Chapter 1. A look at Affective Computing

Affective computing (AC) is an interdisciplinary field joining artificial intelligence, neuroscience, psychology, cognitive science, robotics and more to develop systems capable of recognizing, understanding, generating and expressing the human emotions and reaction of them. Prof. Picard (1995) coined the term defining affective computing as "computing that relates to, arises from, or influences emotions" [1]. Offering new abilities to the computers, in the sense of recognizing, interpreting, processing and manifesting emotions, the human-computer interaction acquires a natural valence [42].

The field of AC aims to develop human emotion-aware technologies and integrate them in domains such as formal and informal learning, healthcare, gaming, robotics, entertainment, virtual reality, marketing, recommender systems, etc. [43], [2]. Furthermore, AC offers the possibility of understanding psychological phenomena and the human behaviour behind the emotion, facilitating the development of software appropriate to the people's needs. Considering the aspect of peoples' affective states in interactions with machines, the human-machine interaction can be viewed as a part of AC.

Even the terms affect, emotion, feeling are often used interchangeably, however they are distinct as highlighted in [44]. An affect is defined as "a non-conscious experience in intensity", feeling as "a sensation that has been checked against previous experiences and labelled", and an emotion as "the projection/display of a feeling. Unlike feelings, the display of emotion can be either genuine or feigned." Niven (2013) defines affect as "the collective term for describing feeling states like emotions and moods" [45]. Emotions and moods are a subcategory of the superordinate category, affect. Also, Niven specifies that emotions are "experiences that are elicited in response to specific external stimuli". There is another term, namely sentiment, which is used in relation to emotion, and we can find the syntagms the sentiment analysis and the emotion recognition used interchangeably. Resuming, AC deals with human emotion, sentiment, and feelings in the computer science field [46]. In this thesis, we accept interchangeably all terms emotion, affect, and sentiment in the context of development of the technologies.

In 2004, affective learning (AL) appeared as a new research study to include in learning situations the technologies, which "help elicit, sense, measure, communicate, understand, reflect upon, and respond to emotions" [2]. Affect-aware learning technologies (AALTs) are introduced to capture the ability of educational technologies to sense the emotion and act

with emotion [3]. In the view of Pekrun (2014), the classroom is "a space of emotions" in which learners' emotions and the learning process influence each other [47]. Moreover, teachers 'emotions affect the whole learning classroom being contributor to the students' cognition, motivation, and outcomes [48]. So, AALTs can be employed in all types of learning activities: shallow, complex and procedural. Mello&Graesser (2015) emphasize that affect influences all three types of activities, particularly the complex one [3]. Considering the above definitions, affective learning can be seen as sub-field of affective computing. Regarding the technologies, AC comprise two major topics: emotion detection/recognition and emotion expression and generation by the machines.

Emotion-aware technologies come with major ethical concerns, which must be addressed. Cowie (2015) argues linking AC to ethics at multiple levels: data, codes, deployments, human subjects in the research, the interactions of humans with technology, usages, perceptions [49]. Six ethical themes directly related to AC are required to be investigated to establish a set of minimal ethical rules: "beneficence, deception, respect for autonomy, certifying competence, portraying humans, application-specific concerns". In the recently adopted AI Act (2024), AI systems for emotion recognition are classified as high-risk systems and are the subjects to specific requirements [4]. Also, the traceability and explainability of the AI systems are demanded to ensure an ethical AI.

Chapter 2. Data for Automated Emotion Recognition

This chapter aims to introduce the emotions models and data used in Automated Emotion Recognition (AER) systems.

Findings

- Our contribution consists in two datasets containing physiological and EEG data acquired in laboratory. The data was used in the research published by the author of this thesis as co-author in [5], [6], [7], [9], [10], [11], [12].

Emotion models

The emotional state of a person cannot be measured directly, in the sense that there is no instrument by which the emotional state can be assessed precisely. A person can manifest involuntarily or not, an emotional expression, which Mandler called a "symptom" [50]. The symptoms are observed, and the humans sense the emotional state of the individual. These symptoms can be easily "seen" by a person from the facial expressions, body gestures, voice tone, eye gaze. In a conversation, the words express the emotional state of a person. With the advancement of technology and the emergence of low-cost wearable sensors, the physiological signals such as heart rate, breathing rate or signals generated from electrodermal activity or signals from the electrical brain activity of individuals can be captured and measured, and the emotions can be assessed. People express and recognize emotions from the multimodal information and more researchers have focused on developing AER systems using not only unimodal but also multimodal information [46], [52], [53]. In Figure 2.1, we highlight the data used in AER-s presented in this thesis.

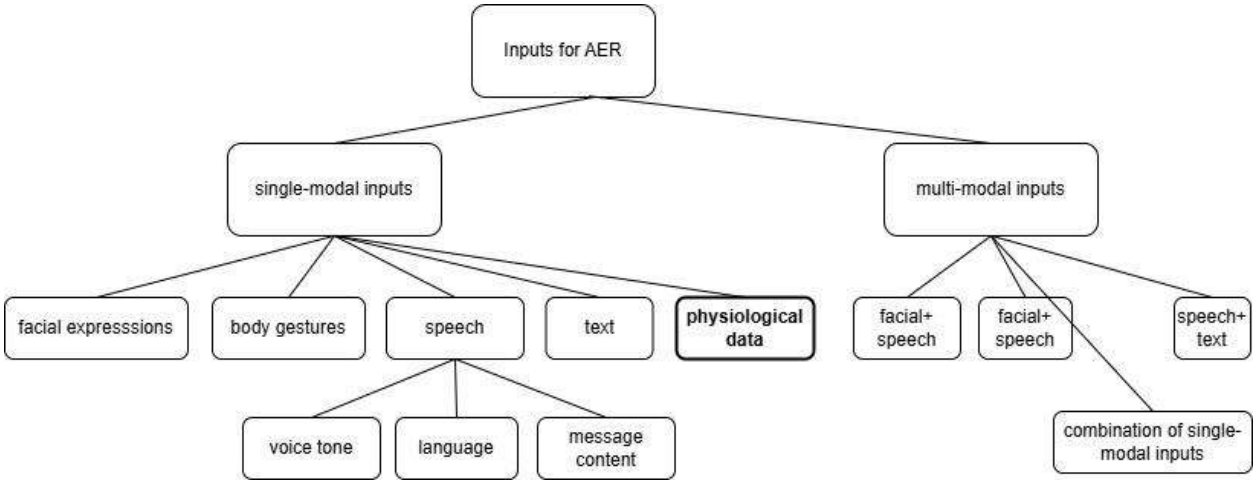

Figure 2.1. Inputs for AER systems

There are three main models for representing emotional states: the discrete model, the componential model, and the dimensional model.

Basic emotions are those that are considered to be universally expressed and recognized by people around the world [54]. A discrete emotion is a category of emotions, for example fear. One of the well-known modern theories of emotions is the Ekman's theory, who initially identified six basic (discrete) emotions: "sadness, happiness, disgust, anger, fear, and surprise" [19]. There are theories that consider other sets of basic emotions, seen as fundamental emotions. Also, different names are used for the same emotion. Izard (1977) launched the differential theory of emotions and proposed ten basic emotions: "interest, joy, surprise, sadness, anger, disgust, contempt, fear, shame, and guilt" [55]. Later, Ekman (1999) identified the attributes that distinguish the basic emotions from each other and considered 15 basic emotions: "amusement, anger, contempt, contentment, disgust, embarrassment, excitement, fear, guilt, pride in achievement, relief, sadness/distress, satisfaction, sensory pleasure and shame" [56]. The complex emotions can be described combining more discrete emotions.

Plutchik's componential model (1980) proposes eight basic emotions and other complex emotions obtained by combining them [57]. The fundamental emotions considered are happiness, trust, fear, surprise, sadness, anticipation, anger and disgust. A combination of two fundamental emotions defines a dyad, for example happiness plus trust defines love. There are also triads, which are obtained by combining three fundamental emotions. The model known as the Wheel/Circle of Emotions is visualized through eight sectors, in which the colour shade expresses the intensity of the emotion.

Cohen (2005) argues that empirical evidence does not support the basic emotion framework because basic emotions cannot be identified by specific functions [58].

The dimensional model proposes the description of an emotional state through the perspective of two or three dimensions: valence and arousal, to which a third dimension, dominance, was added. In Russell's circumplex model (1980), an affective state is described through the values of two dimensions: valence and arousal [59]. Valence (pleasure) reflects positive or negative emotional states, while arousal reflects the level of activation. Arousal is represented vertically, valence horizontally, their intersection is defined by an average activation level and neutral valence. Thus, four quadrants are defined:

- quadrant I – positive valence + high arousal.
- quadrant II – negative valence + high arousal.
- quadrant III – negative valence + low arousal.

- quadrant IV – positive valence + low arousal.

Twenty-eight affective states are represented in space valence-arousal, as follows:

- happy, delighted, excited, astonished, aroused in quadrant I;
- tense, alarmed, angry, afraid, annoyed, distressed, frustrated in quadrant II;
- miserable, sad, gloomy, depressed, bored, droopy, tired in quadrant III;
- sleepy, calm, relaxed, satisfied, at ease, serene, glad, pleased in quadrant IV.

To make a distinction between some emotional states whose points in the 2d space are very close (for example, fear and anger) a third dimension, dominance, was introduced to expresses the degree to which individuals can control emotions.

The PAD (Pleasure-Arousal-Dominance) or VAD (Valence-Arousal-Dominance) model uses three dimensions: pleasure (valence), arousal, and dominance; and was proposed by Mehrabian and Russel (1974, 1977) [60], [61] and later by Mehrabian (1995, 1996) [62], [63]. A correspondence between the VAD model and the discrete model of emotions has been proposed by Russel&Mehrabian (1977) in [61] (Table 2.1). The values for valence, arousal, and dominance are in the interval [-1.1].

Table 2.1. The correspondence between VAD model and the discrete model [61]

	Valence	Arousal	Dominance
anger	-0,43	0.67	0,34
joy	0,76	0,48	0,35
surprise	0,40	0,67	-0,13
disgust	-0,60	0,35	0,11
fear	**-0,64**	**0,60**	**-0,43**
sadness	-0,63	0,27	-0,33

Buechel&Hahn (2016) provide a visual description of the basic emotions in the VAD cube (Figure 2.2) [64].

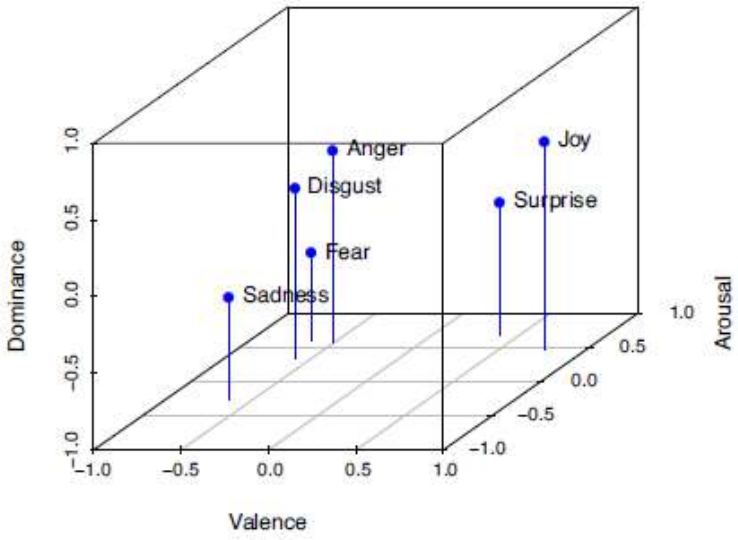

Figure 2.2. Representation of the six basic emotions in VAD cube [64]

In this thesis, in development of our VRET systems we deal with automated recognition for fear, described through low valence, high arousal and low dominance.

For affective learning, we investigated and recognized the emotions often associated with learning process: boredom, confusion, frustration, curiosity, excitement, concentration, anxiety according to Mello&Graesser (2015) [3]. We used PAD model to express each emotion tacking into account the mean PAD values provided in [61] (Table 2.2.).

Table 2.2. Mean PAD values for seven emotions [61]

Emotion	Pleasure	Arousal	Dominance
boredom	-0.65	-0.62	-0.33
confusion	-0.53	0.27	-0.32
frustration	-0.64	0.52	-0.35
curiosity	0.22	0.62	-0.01
excitement	0.62	0.75	0.38
concentration	0.42	0.28	0.39
anxiety	0.01	0.59	-0.15

The mean values for pleasure, arousal, and dominance of the boredom, confusion, frustration, curiosity, excitement, concentration, anxiety are expressed in Figure 2.3 [25].

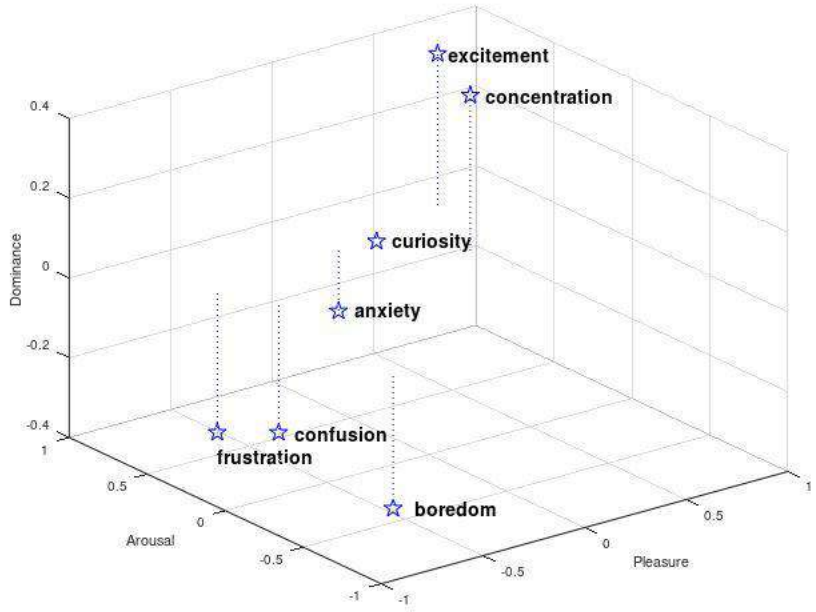

Figure 2.3. PAD representations for seven emotions [25]

Biophysical data

The emotions can be detected based on various signals captured from facial expressions, body language, tone of voice, eye gaze, physiological and EEG recorders and so on. In our research, we used only biophysical data being the most relevant for emotion recognition.

Electroencephalography (EEG) and physiological data

Electroencephalography is a non-invasive technique which records the brain's electrical activity using a special device, called encephalogram. This device has 8, 16 or 32 pairs of electrodes made on a special metal which record the electrical potential at the scalp level The EEG frequency varies from 1 to 100Hz. In fact, an EEG signal is a composite signal with five relevant sub-bands: delta, theta, alpha, beta and gamma. In the Table 2.3, there are described type of waves and the related states [65], [66], [67].

Table 2.3. Type of waves (EEG signals)

Type of waves*	
Delta waves (0.5 - 3 Hz)	psychosomatic relaxation states deep sleep empathy and intuition

Theta waves (4 – 8 Hz)	low brain activities, light sleep, dreams, imagination low related to emotions
Alpha waves (8 – 13 Hz)	normal wakeful state reflection, "gateway to our creativity" medium related to emotions
Beta waves (13 – 30 Hz)	active thinking, high concentration consciousness, brain activities and motor behaviors high related to emotions
Gamma waves (>30 Hz)	memory and language processing high related to emotions

*The ends of the intervals vary by 1-2 Hz in various studies

In short, EEG measures electrical activity during synaptic excitation of neurons.

Physiological data

Galvanic Skin Response (GSR) represents a measure of the skin conductance, and it can be decomposed into two components (Figure 2.4):

- Skin Conductance Level (SCL) – background tonic with mean value 2-20 μS.
- Skin Conductance Response (SCR) with non-specific SCR and event-related SCR – rapid phasic components with mean value 0.1-1.3 μS.

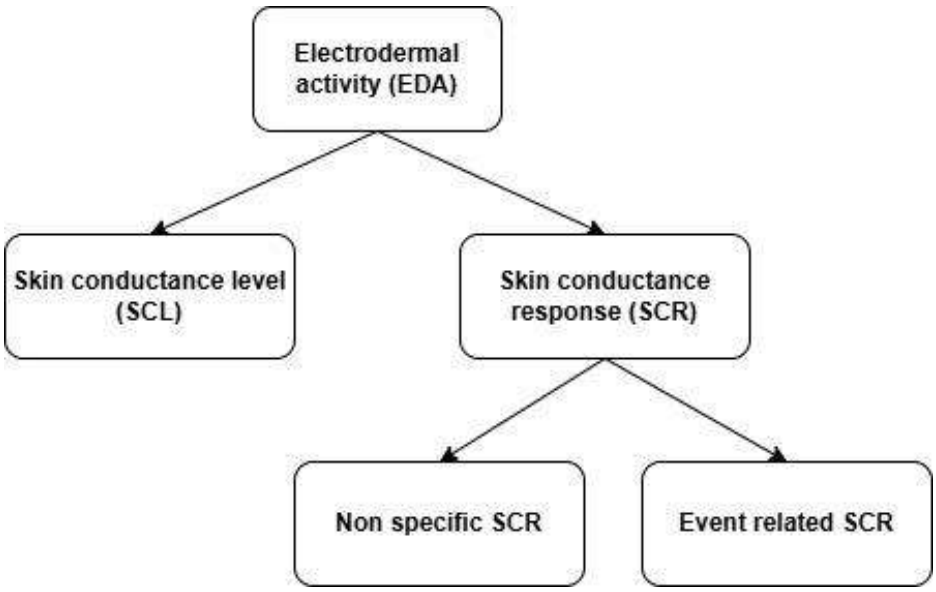

Figure 2.4. Diagram of Electrodermal signals

To determine GSR values, there are used two electrodes placed on the middle and index fingers.
Electromyography (EMG) records the electrical activity of the muscles, respiration rate and body temperature.
Heart Rate (HR) measures the number of heart beats in one minute (bpm).
Heart Rate Variability (HRV) measures the variations of time between heart beats.
Respiration Rate (RR) measures the number of breaths per minute.

Biophysical datasets for AER

Considering that the most reliable data in association with emotions is biophysical data, we mainly used three datasets, both containing physiological and EEG data. Two datasets (Dataset 1 and Dataset 2) were built by our team through a series of experiments performed in laboratory [5], [6], [7], [9], [10], [11], [12].
Dataset 1 – We captured EEG, EDA and HR signals from 4 subjects in both virtual and in real-environment. The subjects were exposed to different height levels, and we acquired 63 trials per subjects obtaining 25000 entries for each subject.
For EEG signals (channels: FP1, FP2, FC5, FC1, FC2, FC6, T7, C3, C4, T8, P3, P1, P2, P4, O1 and O2) we used Acticap Xpress Bundle device with 16 dry electrodes. EDA and HR signals were captured with the GSR unit of a Shimmers Multi-Sensory device. Data captures and their usage are presented in Chapter 3, section Development of the Adaptive VRET Systems with Emotion Recognition.
Dataset 2 – We acquired GSR, HR and RR from 5 subjects in two phases. In the first phase, we measured GSR, HR and RR while the subjects were performing the following actions deep breath, head movement to the left, head movement to the right, head movement up, head movement down, click with the right hand on the HTC Vive controller, right hand raise. In the second phase, the participant played a VR-based game, and we performed the same measurements. For HR and GSR measurements, we used the Shimmer3 GSR+ Unit (https://www.shimmersensing.com/product/shimmer3-gsr-unit/). The Respiration Rate was calculated according to the distance between two HTC Vive trackers. Data captures and their usage are presented in Chapter 3, section Challenges for Integrating AER in VRET system.

Dataset 3 - DEAP (Database for Emotion Analysis using Physiological signals) dataset [13].

The third dataset is a benchmark one, namely DEAP (Database for Emotion Analysis using Physiological signals) dataset [13]. DEAP dataset is an intensively used database in AER systems. It contains electroencephalography (EEG) signals and peripheral physiological signals (galvanic skin response - GSR, respiration amplitude, skin temperature, electrocardiogram, blood volume by plethysmograph, electromyograms of Zygomaticus and Trapezius muscles, and electrooculogram - EOG), acquired from 32 subjects.

The distribution of the subjects was 50% male and 50% female and the mean age 26.9. The participants watched 40 music videos and rated valence, arousal, dominance, like/dislike with a float value from 1 to 9 and the familiarity with and integer from 1 to 5.

Using AgCl 32 electrodes positioned according 10/20 system, EEG signals at a sampling rate of 512 Hz were recorded: FP1, AF3, F3, F7, FC5, FC1, C3, T7, CP5, CP1, P3, P7, PO3, O1, Oz, Pz, FP2, AF4, Fz, F4, F8, FC6, FC2, Cz, C4, T8, CP6, CP2, P4, P8, CP6, CP2, P4, P8, PO4, O2. The peripheral physiological signals GSR were recorded using EXG sensors face, EXG sensors trapezius, respiration belt, and left physiological sensors. More details about DEAP dataset can be found in [13].

More physiological datasets for emotion recognition are available for research. A selection of them is presented in Table 2.4.

Table 2.4. Physiological datasets for emotion recognition

DEAP - Database for Emotion Analysis using Physiological signals	[13] http://www.eecs.qmul.ac.uk/mmv/datasets/deap/readme.html
MAHNOB-HCI	[68], https://mahnob-db.eu/
eNTERFACE06_EMOBRAIN	[69]
SEED	[70], https://bcmi.sjtu.edu.cn/home/seed/
BioVid Emo DB	[71]
NeuroMarketing	[72]

Chapter 3. Emotion-Aware Virtual Reality Exposure Therapy Systems for Phobia Treatment

This chapter presents the results of the research started in 2018 in the project entitled System for Ameliorating Phobias based on Exposure to Virtual Reality (SAFE-VR)[1] coordinated by Oana Mitruț (University Politehnica of Bucharest), on which the author of this thesis collaborated. The project goal was to perform a proof of concept of the emotion-aware Virtual Reality-based system for treating phobia. We had to recognize the users' emotions by monitoring physiological signals using various machine learning models. The system had to automatically and in real time adapt to the users' emotions. To achieve the goal, more machine learning models for emotion recognition were developed and integrated into two VR-based video games. The results of the research related to AER in VR-based phobia treatment are published by the author of this thesis as co-author in papers [5], [6], [7], [8], [9], [10], [11], [12], [14], [15], [16].

Findings

- For DEAP dataset, we obtained two Random Forest classifiers for fear according to the two-level scale (0- the two-level scale with 0 – absence of fear and 1 – presence of fear) and the four-level scale (0 – absence of fear, 1 – presence of low fear, 2 – presence of medium fear and 3 presence of high fear). We extracted the Power Spectral Density (PSD) of all 32 EEG channels in the alpha, beta and theta frequency ranges and computed the mean values for alpha, beta and theta PSD in the pre-frontal (FP), AF (between FP and F), frontal (F), FC (between F and C), central (C), temporal (T), P (parietal), CP (between C and P), O (occipital) and PO (between P and O) sides of the brain. We obtained a dataset with 30 EEG features and 8 physiological features (hEOG, vEOG, zEMG, tEMG, GSR, Respiration, PPG and temperature). The resulting dataset fed RF classifiers with no feature selection, and we obtained the following performance accuracy – 93.13%, F1 score – 93.11% in the case of 2-level scale and accuracy – 85.74%, F1 score – 85.33% in the case of 4-level scale.
- For DEAP dataset, we developed more ML models for recognizing 6 discrete emotions: sadness, happiness, disgust, anger, fear, and surprise.

[1] http://nets.cs.pub.ro/~safe_vr/contact.html

- For DEAP dataset (only EDA, HRV features), we proposed a feature extraction protocol to segment each trial from DEAP both in three non-overlapping windows and five overlapping windows. We tested more ML models and obtained the most performant according to ROC AUC score:
    - for the non-overlapping dataset: PCA reduction + SVM – 93.5%.
    - for the overlapping dataset: GBT – 91.7%,
- We developed an emotion-aware VRET system based on a video game for acrophobia treatment using two types of deep neural networks: one for fear level classification and one for determining the next level of the game, according to the desired level of fear. We tested each paradigm, 2, 4, 11 - choices scale for fear level assessment, and obtained the most performant:
    - Three AER models for fear level classification with maximum cross-validation accuracies 95.51% – 2-choices scale; 90.49% - 4-choices scale; 85.09% – 11 choices scale.
    - Three models for game level prediction with maximum cross-validation accuracies 98.72 – 2-choices scale; 98.67% - 4-choices scale; 98.75% – 11 choices scale.

  We used Alpha, beta, theta, theta/beta sub-channels from the channels: FP1, FP2, FC5, FC1, FC2, FC6, T7, C3, C4, T8, P3, P1, P2, P4, O1, O2 with the following values: Alpha FP2- Alpha FP1; (Alpha FC2 + Alpha FC6) – (Alpha FC5 +Alpha FC1); (Alpha FP2 + Alpha FC2 + Alpha FC6) – (Alpha FP1 + Alpha FC5 + Alpha FC1); (Beta FP2 + Beta FC2 + Beta FC6) – (baseline Beta FP2 + baseline Beta FC2 + baseline Beta FC6); (theta/beta FC6+ theta/beta FC2 + theta/beta FP2) – (baseline theta/beta FC6+ baseline theta/beta FC2 + baseline theta/beta FP2) and GSR and HR values.
- We defined a ML-based decision support with kNN, SVM, RF, LDA, DNN classifiers for both predictions level of fear and the game level that should be played next. We considered two modalities to calculate the accuracies of the classifiers: a user-dependent and a user-independent. With respect to the fear classifier, the highest cross validation accuracy (over 98%) was obtained by using either the kNN or RF algorithms, for both the player-independent and player-dependent modalities. Moreover, for the game level classifier, we observed the high performance with the RF algorithm.
- We proved the necessity of the human-in-the-loop strategy in developing VRET systems with ML models to ensure the patient's safety through a strong collaboration

between patient, therapist and machine. Data for the ML models need to be acquired in the application-specific context, the therapist must supervise the entire process, including in the VRET systems development and treatment.

- We defined a Human-Centered VRET System Design Methodology using layers-based analysis from the Human-Centered Distributed Information Design (HCDID) model [73] and from the Needs and Aspirations for application in a Design and Innovation (NADI) model [74].
- We proposed a holonic architecture for VRET systems.
- We provided a method to identify and remove artefacts in biophysical data during a VR-based game playing.
- We designed and tested a protocol for the acquisition and processing of biophysical signals in VR environments, particularly in phobia treatment.

The Importance of the topic

Mental health disorders affect a significant percentage of the world population. Mainly, three types of phobias exist: agoraphobia (fear of public places), social phobia (fear of meeting or relating to other people) and specific phobias (anxieties produced by objects and situations). The manifestations of the phobias consist in increasing the heart rate, heavy sweating, rapid breathing, hand kneading, even loss of consciousness and many more symptoms as the results of the intensification of the activity in the nervous and sympathetic systems.

In our research, we considered the acrophobia (fear of height), a type of specific phobia defined in [75] as "a marked and excessive fear or anxiety that consistently occurs upon exposure or anticipation of exposure to one or more specific objects or situations (e.g., proximity to certain animals, flying, heights, closed spaces, sight of blood or injury) that is out of proportion to actual danger. The phobic objects or situations are avoided or else endured with intense fear or anxiety. Symptoms persist for at least several months and are sufficiently severe to result in significant distress or significant impairment in personal, family, social, educational, occupational, or other important areas of functioning".

According to [17] a percentage between 5% and 10% of the world population is affected by a specific phobia, and acrophobia (fear of height) has a high incidence affecting 1 in 20 individuals. In a report published by the Institute for Health Metrics and Evaluation, we found that mental disorders increased from rank 9 in 1990 to rank 6 in 2021 [76]. The most

affected regions are Portugal with 19936 cases per 100000 followed by Iran, and Lebanon. High incidence of mental disorders can be found also in Australia, Spain, France, Ireland, US, Brazil, New Zeeland. In Romania are reported 12217 cases per 100000. In total 13.9% of the population experienced a mental disorder in 2021 and 71% of these diseases could be avoided if the population had access to optimal treatment.

One of the successful treatments for specific phobia is Cognitive-Behavioural Therapy (CBT) - including two methods: cognitive and exposure (behavioural) therapy. Exposure therapy consists in gradually exposing people to anxiety, eliciting objects or situations with the goal of changing the people's responses towards the fear. The therapy must be supervised and guided by a specialist. CBT can be performed in-vivo or virtually in a controlled environment. Through this procedure, patients learn to acknowledge and control their situations.

VR technology was used in exposure therapy since the end of 1990, when a pilot experiment was conducted for the treatment of specific phobia (fear of flying, heights, public speaking, and fear of being in certain situations) [18]. Virtual Reality Exposure Therapy (VRET) has been successful, with more and more people preferring this type of therapy over in-vivo exposure therapy. We have noted the experiments performed by Garcia-Palacios et al. (2001), in which more than 80% of the subjects involved in the experiments chose VRET instead of in-vivo [77].

In our research we performed more experiments in 2018 and 2019 regarding usage of the VR-based technology in phobia treatment having automated emotion recognition. We developed two video games adapting to the users' emotions. With machine learning-based algorithms we predicted the level of exposure in various situations for the subjects involved in the experiments in order to mitigate the fear of heights. Our solution to phobia treatment was seen as an AI-assistant, keeping the human in the loop of the process.

AER models based on DEAP datasets

In this section we present more machine learning models developed for fear's levels recognition based on DEAP dataset. The models are published in papers [14], [15]. We extended the work presented in [14] including all six basic emotions from Ekman's theory (sadness, happiness, disgust, anger, fear, and surprise) in [16].

The fear emotion is described in VAD model as low valence, high arousal and low dominance. In DEAP datasets, the evaluations of the valence, arousal, and dominance attributes are in the range [1,9]. So, we used in [14] two modalities for fear' intensity assessment:

- the two-level scale with 0 – absence of fear and 1 – presence of fear.
- four-level scale with 0 – absence of fear, 1 – presence of low fear, 2 – presence of medium fear and 3 - presence of high fear.

The divisions of the range [1, 9] in the two paradigms are shown in Table 3.1 and Table 3.2.

Table 3.1. The modality 2 - level scale for fear

Label	Valence	Arousal	Dominance
0 - no fear	(5; 9]	[1; 5)	[5; 9]
1 - fear	[1; 5]	[5; 9]	[1; 5)

Table 3.2. The modality 4 -level scale for fear

Label	Valence	Arousal	Dominance
0 - no fear	[7; 9]	[1; 3)	[7; 9]
1 - low Fear	[5; 7)	[3; 5)	[5; 7)
2 - medium fear	[3; 5)	[5; 7)	[3; 5)
3 - high fear	[1;3)	[7;9)	[1;3)

The steps followed to obtain the two classifiers for fear according 2-scale and 4-scale paradigms, are presented in Figure 3.1.

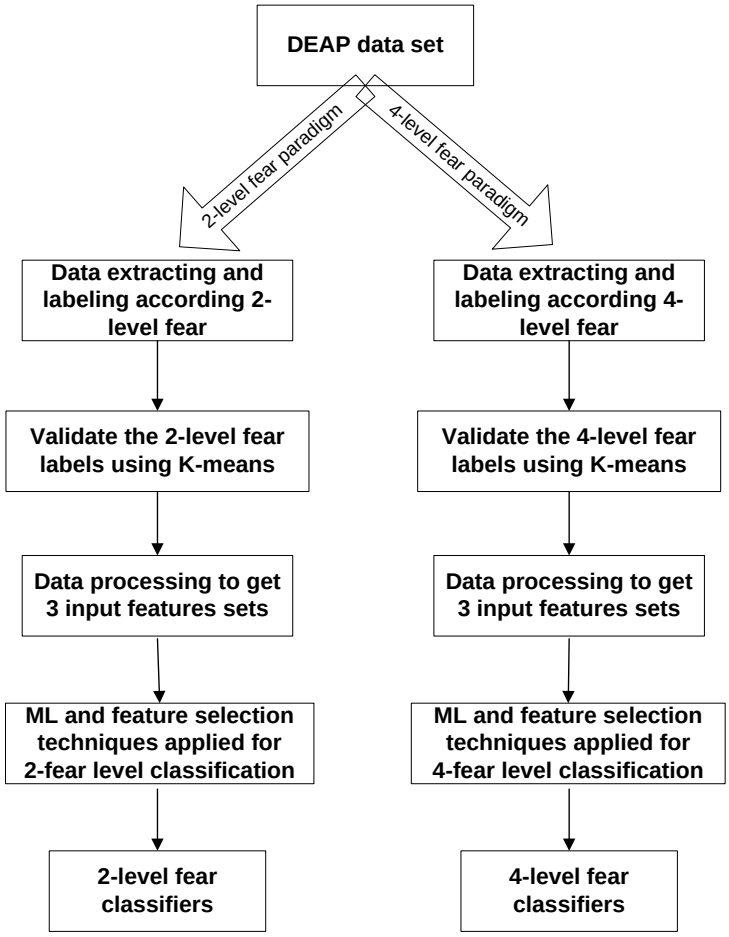

Figure 3.1. The schema for obtaining fear classifiers [14]

We labelled data for both modalities according to the values presented in Table 3.1 and Table 3.2. The proposed labelling was validated with K-Means algorithm, obtaining the average accuracy 87%.

The number of samples for two-level scale was initially: 198 entries – no fear condition and 174 entries - fear condition and for four-level scale: 7 entries - no fear condition; 60 entries – low fear; 42 – medium fear condition; 35 – high fear condition. In DEAP, the duration of the data recording is 60s, so we decided to divide each segment of 60 seconds into in 12 segments, each 5 seconds long. Thus, we obtained in total 4464 entries for the two-level modality and 1728 entries for the four-level modality.

After that, we built five inputs sets for each paradigm, 2-level and 4-level scale for fear:

- Set 1 - Raw EEG values from 32-channel and physiological data - hEOG, vEOG, zEMG, tEMG, GSR, Respiration, PPG and temperature). There were obtained 40 features (32+8).
- Set 2 - Power Spectral Density (PSD) of all 32 EEG channels in the alpha, beta and theta frequency ranges. Then, there were calculated the mean values for alpha, beta and theta PSD in the pre-frontal (FP), AF (between FP and F), frontal (F), FC (between F and C), central (C), temporal (T), P (parietal), CP (between C and P), O (occipital) and PO (between P and O) sides of the brain. So, there were obtained a dataset with 30 EEG inputs (3*10) and 8 physiological recordings (hEOG, vEOG, zEMG, tEMG, GSR, Respiration, PPG and temperature).
- Set 3 - Petrosian Fractal Dimension of 32 EEG channels and the physiological recordings (hEOG, vEOG, zEMG, tEMG, GSR, Respiration, PPG and temperature).
- Set 4 - Higuchi Fractal Dimension of 32 EEG channels and the physiological recordings (hEOG, vEOG, zEMG, tEMG, GSR, Respiration, PPG and temperature).
- Set 5 Approximate Entropy for each of the 32 EEG channels and the physiological recordings (hEOG, vEOG, zEMG, tEMG, GSR, Respiration, PPG and temperature).

In the case of sets 3, 4, 5 there were obtained 40 features (32+8) for each set.

We built more machine learning models: four deep neural networks and four ML models - Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Random Forest (RF) and k-Nearest Neighbors (kNN) and applied more feature selection - Principal Component Analysis (PCA), Sequential Feature Selector (SFS), Fisher selection for each input set (1-5).

DNN1 model consists in – one input layer, three hidden layers with 300 neurons per layer, one output layer with binary crossentropy loss function for two-level scale (with two possible outputs, 0 or 1, where 0 stands for no fear and 1 for fear) and the categorical cross-

entropy loss function for the four-level scale (one-hot encoding for each class). Further, it was used an Adam gradient descent optimization algorithm, and for the activation function we chose RELU. The data were standardized. For training, there had been set 1000 epochs and a batch size of 20. The cross-validation was performed by using the k-fold method with k = 10 splits. DNN2 model has three hidden layers with 150 neurons per layer; DNN3 model – six hidden layers with 300 neurons per layer; DNN4 model – six hidden layers with 150 neurons per layer. Also, we used SVM, RDF, LDA and kNN models. For each model, we performed 10 times cross-validation, without feature selection and with feature selection (Fisher, PCA, Sequential Feature Selection). The division of the dataset was: 70% for training and 30% for test. The performances obtained for each case are presented in Tables 3.3-3.7.

Table 3.3. The performance of the models for Set 1 (Raw EEG values from 32-channel and 8 peripheral signals) [14]

Classifier	2-level scale		4-level scale	
	F1 score (%)	Accuracy (%)	F1 score (%)	Accuracy (%)
No feature selection				
DNN1	70.59	70.95	58.78	59.84
DNN2	67	67.34	34.16	45.78
DNN3	71.91	71.95	47.69	51.16
DNN4	69.17	69.27	24.51	41.67
SCN	75.15	76	48.35	49
SVM	73.5	74	64.65	66.09
RF	85.63	86	**83.85**	**84.01**
LDA	60.19	61	55.77	56.65
kNN	**86.81**	**87**	82.52	82.66
Fisher selection				
SVM	69.95	70.90	60.03	62.24
RF	84.41	84.55	77.81	78.03
LDA	53.29	55.90	46.46	48.17
kNN	**84.70**	**84.70**	**79.22**	**79.19**
PCA selection				
SVM	74.20	74.87	69.93	70.83
RF	82.30	82.45	78.32	78.54
LDA	56.87	58.27	46.43	49.36

kNN	**85.54**	**85.53**	**81.67**	**81.75**
SFS selection				
SVM	**60**	**60**	**48**	**48**
RF	57	57	43	43
LDA	59	59	**48**	**48**
kNN	57	57	46	46

Table 3.4. The performance of the models for Set 2 (30 alpha, beta and theta PSDs and 8 peripheral signals) [14]

Classifier	2-level scale		4-level scale	
	F1 score (%)	Accuracy (%)	F1 score (%)	Accuracy (%)
No feature selection				
DNN1	81.99	81.99	67.46	68.98
DNN2	78.16	78.14	55.92	58.85
DNN3	82.21	82.26	57.70	60.94
DNN4	79.14	79.12	30.13	43.63
SCN	75.12	75.5	51.2	51.5
SVM	83.15	83.13	83.46	84.01
RF	**93.11**	**93.13**	**85.33**	**85.74**
LDA	70.46	70.52	60.98	61.46
kNN	85.84	85.82	82.94	83.24
Fisher selection				
SVM	78.15	78.13	76.79	77.07
RF	**85.68**	**85.75**	80.28	80.54
LDA	64.90	65.37	51.66	52.99
kNN	81.05	81.04	**82.60**	**82.66**
PCA selection				
SVM	85.82	85.81	82.52	82.85
RF	84.75	84.83	81.08	81.27
LDA	65.94	66.13	54.93	55.39
kNN	**87.45**	**87.44**	**82.58**	**82.77**
SFS selection				
SVM	71	71	56	56

RF	**81**	**81**	**68**	**68**
LDA	64	64	48	48
kNN	78	78	61	61

Table 3.5. The performance of the models for Set 3 (32 Petrosian Fractal Dimensions and 8 peripheral signals) [14]

Classifier	2-level scale		4-level scale	
	F1 score (%)	Accuracy (%)	F1 score (%)	Accuracy (%)
No feature selection				
DNN1	80.90	80.91	62.65	64.35
DNN2	77.65	77.64	39.56	48.96
DNN3	80.08	80.17	49.11	56.60
DNN4	76.47	76.50	24.51	41.67
SCN	78.6	78.75	47.34	48.15
SVM	81.57	81.57	**82.01**	**82.47**
RF	**87.33**	**87.54**	68.57	69.75
LDA	62.94	62.99	52.09	52.79
kNN	83.23	83.21	77.09	77.26
Fisher selection				
SVM	81.49	81.49	72.96	73.60
RF	**87.33**	**87.54**	68.57	69.75
LDA	62.94	62.99	52.09	52.79
kNN	83.23	83.21	**77.09**	**77.26**
PCA selection				
SVM	81.77	81.78	79.05	79.69
RF	79.35	79.62	70.68	71.46
LDA	64.41	64.58	63.98	64.32
kNN	**85.37**	**85.36**	**83.50**	**83.64**
SFS selection				
SVM	71	71	56	56
RF	**80**	**80**	**66**	**66**
LDA	67	67	51	51
kNN	78	78	61	61

Table 3.6. The performance of the models for Set 4 (32 Higuchi Fractal Dimensions and 8 peripheral signals) [14]

Classifier	2-level scale		4-level scale	
	F1 score (%)	Accuracy (%)	F1 score (%)	Accuracy (%)
No feature selection				
DNN1	81.40	81.41	59.89	62.67
DNN2	77.01	76.99	36.96	47.74
DNN3	81.09	81.14	49.11	57.12
DNN4	78.51	78.52	24.51	41.67
SCN	77.15	78.5	45.25	46.20
SVM	81.64	81.64	80.85	81.70
RF	**89.96**	**90.07**	**82.59**	**83.24**
LDA	69.09	69.10	64.96	65.32
kNN	83.38	83.36	80.52	80.73
Fisher selection				
SVM	80.75	80.75	71.58	72.45
RF	**88.96**	**89.10**	**77.32**	**72.83**
LDA	66.59	66.87	55.35	56.45
kNN	83	82.99	75.61	75.92
PCA selection				
SVM	82.16	82.16	77.79	78.63
RF	82.26	82.41	78.55	78.90
LDA	69.06	69.16	61.38	61.89
kNN	**84.31**	**84.30**	**81.01**	**81.21**
SFS selection				
SVM	74	74	58	58
RF	**81**	**81**	**68**	**68**
LDA	66	66	48	48
kNN	78	78	61	61

Table 3.7. The performance of the models for Set 5 (32 Approximate Entropies and 8 peripheral signals) [14]

Classifier	2-level scale		4-level scale	
	F1 score (%)	Accuracy (%)	F1 score (%)	Accuracy (%)
No feature selection				
DNN1	79.95	80.17	57.96	61.86
DNN2	79.02	79.21	48.20	54.34
DNN3	80.12	80.58	52.05	59.95
DNN4	79.96	80.40	27.55	41.90
SCN	80.20	80.40	51.25	51.30
SVM	74.71	74.70	57.85	62.43
RF	**89.65**	**89.78**	**80.78**	**81.70**
LDA	62.81	62.91	49.12	52.22
kNN	84.54	84.55	71.15	71.87
Fisher selection				
SVM	75.75	75.75	60	64.35
RF	**89.51**	**89.63**	**80.47**	**81.50**
LDA	58.46	59.93	46.09	50.67
kNN	84.96	85	78.82	79.38
PCA selection				
SVM	77.63	77.78	64.63	68
RF	82.40	82.54	73.49	74.03
LDA	62.75	63.05	48.80	52.45
kNN	**84.93**	**84.95**	**75.97**	**76.61**
SFS selection				
SVM	72	72	55	55
RF	**80**	**80**	**63**	**63**
LDA	64	64	48	48
kNN	78	78	62	62

Analysing the results we concluded that high performances for predictions were achieved using Set 2 data, with RF and no feature selection (accuracy – 93.13%, F1 score – 93.11%) in the case of 2-level scale and similarly using Set 2 data, RF and no feature selection (accuracy – 85.74%, F1 score – 85.33%) in the case of 4-level scale.

The most relevant features ranked by Fisher selection were the raw EEG values of the F4, FC2, CP5 and C3 electrodes, the alpha amplitude in the AF channel, the beta amplitudes in AF and PO, the theta intensities of the AF, P, O and PO electrodes, the Pz Petrosian value, FC5 and C3 for Higuchi, PO3 and O1 Approximate Entropies, and for the peripheral features were PPG, temperature and respiration rate. Also, SFS algorithm ranked as the relevant features: the FP1 and AF3 raw EEGs, the beta amplitudes in the AF and F channels and the theta amplitudes in the C and CP channels respectively, the FC1 and CP1 Petrosian values, the F3, CP1 and P3 Higuchi values and the FP1, AF3 and F7 Approximate Entropies. Related to peripheral features, we obtained that GSR, tEMG and respiration rate were the most relevant.

We extended in [16] the approach used in [14] for recognition (presence/absence) of the six basic emotion, namely sadness, happiness, disgust, anger, fear and surprise identified by Ekman in [19]. We used the values from Tables 2.1. for the correspondence between VAD model and discrete model of emotions. To estimate the intervals for valence, arousal and dominance values in interval [1,9], we considered the ratings low ([1,5)) or high ([5.9]) for valence and arousal and for dominance a narrower interval. As input features, we had EEG (raw values/approximate entropy/Petrosian fractal dimension/Higuchi fractal dimension) and peripheral signals, hEOG, vEOG, zEMG, tEMG, GSR, respiration rate, PPG and temperature. Testing the models developed in [14], we obtained the best classification F1 scores to each emotion according to the values presented in Table 3.8.

Table 3.8. Best F1 scores classification for discrete emotions [16]

	Raw		Petrosian		Higuchi Fractal dimension		Approximate entropy	
	No feature selection	Feature selection	No feature selection	Feature selection	No feature selection	Feature selection	No feature selection	Feature selection
Anger	RF 96.04	kNN Fisher 97.52%	SVM **98.02%**	SVM Fisher 95.05%	SVM **98.02%**	SVM Fisher 94.05%	RF 92.55%	RF Fisher 92.08%
Joy	kNN 91.22%	LDA SFS **100%**	kNN 87.9%	kNN Fisher 85.76%	kNN 87.60%	kNN Fisher 83.75%	RF 86.40%	RF Fisher 80.37%
Surprise	kNN 85.01%	SVM SFS **96%**	kNN 84.75%	kNN Fisher 80.59%	kNN 83.64%	RF Fisher 80.69%	kNN 81.30%	kNN Fisher 82.59%

Disgust	RF **93.63%**	kNN Fisher 89.74%	kNN 95%	SVM Fisher 90%	SVM 91.59%	RF Fisher 90.23%	RF 83.14%	RF Fisher 75.26%
Fear	kNN 90.75%	RF Fisher 80.85%	kNN **89.72%**	kNN Fisher 80.82%	kNN 89.04%	kNN Fisher 83.39%	kNN 80.66%	kNN Fisher 79.45%
Sadness	RF 87.49%	kNN Fisher 85.76%	kNN 90.17%	kNN Fisher 83.29%	SVM **90.8%**	SVM Fisher 86.43%	RF 81.86%	kNN Fisher 76.45%

In subsequent research [15], we considered only GSR and HRV signals from DEAP dataset and extracted 33 types of features for EDA and 7 types of features for HRV, in total 40 type of features. We proposed a feature extraction protocol segmenting each trial both in three non-overlapping windows and five overlapping windows. The pipeline for the process applied in [15] is shown in Figure 3.2.

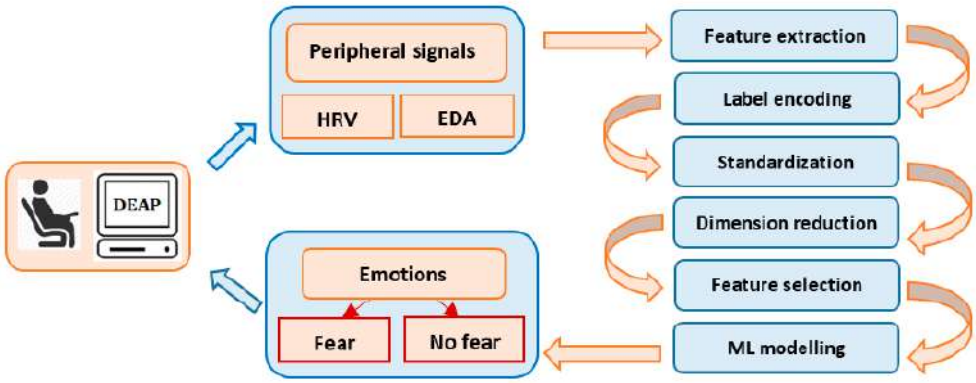

Figure 3.2. The pipeline of the process of automated fear recognition [15]

In the non-overlapping approach, each segment of 60 seconds was divided into three non-overlapping windows of 20 seconds each. So, in this situation we evaluated 120 features (3 segments x 40 features on each segment). In the overlapping approach, each segment was divided into five windows of 20 seconds each and 10 seconds overlapping other segments. Thus, we obtained 40 features for each trial.

From EDA there were extracted more features: time, frequency, events-related and from the plethysmograph. All formulas are detailed in [15].

We obtained the following time-related features:

- mean absolute values of GSR, SCL and SCR signals (related to the evaluation window).
- the ratio between mean absolute values of GSR, SCL and SCR over the evaluation window and over the baseline interval.
- the standard deviations of GSR, SCL and SCR signals over the evaluation window.

- the ratio between standard deviations of GSR, SCL and SCR over the evaluation window and over the baseline interval.
- the Waveform Length of the GSR, SCL and SCR signals.
- the Slope Sign Changes of the GSR, SCL and SCR signals.
- the Willison Amplitudes of the GSR, SCL and SCR signals.

The frequency related features were calculated based on the Power Spectrum Density of the EDA signals:

- the median frequency of the power spectrum.
- the mean frequency of the power spectrum.
- the ration between the signal energy in the [0.2 Hz, 2 Hz] and the [0.01 Hz, 0.2 Hz] frequency intervals.

The events-related features used the results of Continuous Decomposition Analysis:

- the number of electrodermal responses detected inside the window.
- the average amplitude of the electrodermal responses.
- the maximum of the amplitude of the responses inside the window.

Also, seven features were obtained from the plethysmograph:

- the standard deviation of the intervals between successive heartbeats.
- the standard deviation of the difference between successive heart rates inside the window.
- the numbers of successive heartbeat durations that differ by more than 50/20 ms inside the window and the ratio between these numbers and the total number of heart beats in the window.
- the ration between the HR signal energy in the [0.15 Hz, 0.4 Hz] and the [0.04 Hz, 0.15 Hz] frequency intervals.

Finally, we used two datasets, one for the non-overlapping paradigm (120 features) and one for the overlapping paradigm (40 features). For the non-overlapping dataset, there was used 992 sampled labelled with 0 - non fear condition and 166 samples labelled with 1- presence of fear condition. The rule used for labelling data was:

- 1 – presence of fear if $\text{valence} \leq 5 \text{ AND arousal } > 5 \text{ AND dominance} \leq 5$.
- 0 – absence of fear in rest.

Because the classes were imbalanced, the majority class was randomly undersampled and the minority class was oversampled. Finally, the non-overlapping dataset consisted in 1494 observations (830 for non-fear condition; 664 for presence of fear condition). In the case of

the overlapping paradigm, we started from 4960 observations for class 0 and 830 for class 1 and obtained 4150 observations for class 0 and 3320 observations for class 1.

More machine learning models were designed to obtain a binary classifier (absence of fear, and presence of fear) for each paradigm based on Decision Trees – RF (Random Forest) and GBT (Gradient Boosting Tree), k-Nearest Neighbors (kNN), Support Vector Machine (SVM) and shallow and deep artificial neural networks with predefined number of neurons and hidden layers with various optimization techniques. In the case of SVM and KNN algorithms, the data has been scaled.

For SVM we used the C-Support Vector Classification (SVC) function from the scikit-learn library with RBF kernel. To reduce the dimensionality, we applied PCA and for feature selection, we considered the algorithms: XGBoost, Pearson Correlation Coefficient, L1 regularization, RF Classification and Recursive Feature Elimination. We choose the first k relevant features (k= 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110 and 120 for the non-overlapping dataset and 5, 10, 20, 30 and 40 for the overlapping dataset) for each feature selection algorithm. We performed multiple combinations between the values for the penalty C and for the kernel coefficient γ during a five-fold cross-validation on the training set. The values considered were the following: for the penalty C – 0.1; 1; 10; 100; 1000; 10000 and for the kernel coeficient γ – 1/no_features; 1/(no_features*10); 1/(no_features*100); 1/(no_features*1000).

For RF, we used the Gini criterion, each individual tree was trained on different sets of features. The values considered for the hyperparameters were:

- The maximum depth of an individual tree (max_depth) - 3, 10, none.
- The minimum number of samples required for node splitting (min_sample_split) – 10, 30, 50, 70, 90.
- The number of features required for the best node splitting (max_features) - sqrt(no_features) and log2(no_features).
- The number of trees for majority voting (n_estimators) – 100, 300, 500.

As in the case of SVM algorithm, we performed five-fold cross validation.

In the case of GBT, we have set the learning rate at 0.1, the max_depth at 10 and n_estimators at 1000 and, we applied five-fold cross validation.

For kNN, we used the leaf_size in the range 5 to 50, n_neighbors between 3 and 9, p (the power parameters for the Minkowski metric) between 1 and 2. Also, we proceeded as in the case of SVM algorithm and applied PCA and feature selection with: XGBoost, Pearson Correlation Coefficient, L1 regularization, RF Classification. We gave up on the Recursive

Feature Elimination, because it was too computationally expensive. We performed five-fold cross validation.

For neural networks, both shallow and deep, we tested more combinations of the number of neurons on the hidden layer, the number of epochs, the learning rate, and the dropout rate (Table 3.9. in the case of the non-overlapping dataset and Table 3.10. in the case of the overlapping dataset).

Table 3.9. The configurations of the neural networks for the non-overlapping dataset [15]

	No. of hidden layers	No. of neurons on the hidden layer	Dropout	No. of epochs
Configuration 1	-	-	-	200
Configuration 2	-	-	0.5 for the input layer	200
Configuration 3	-	-	0.8 for the input layer	500
Configuration 4	1	80	0.8 for the input layer	500
Configuration 5	1	11	0.8 for the input layer	500

Table 3.10. The configurations of the neural networks for the overlapping dataset [15]

	No. of hidden layers	No. of neurons on the hidden layer	Dropout	No. of epochs
Configuration 1	-	-	-	200
Configuration 2	-	-	0.5 for the input layer	200
Configuration 3	-	-	-	500
Configuration 4	-	-	0.5 for the input layer	500
Configuration 5	1	6	0.5 for the input layer	200
Configuration 6	2	27 and 6	0.5 for the input layer	200
Configuration 7	3	30, 20 and 10	0.5 for the input layer	200

We present the results for models obtained for the two types of datasets: non-overlapping and overlapping datasets (Table 3.11-3.20). We used five metrics to assess the models: accuracy, F1 score, ROC AUC score, sensitivity, and specificity.

Table 3.11. The results for the SVM-based model - the case of the non-overlapping dataset [15]

	SVM	PCA Dimensionality Reduction + SVM	XGBoost Feature Selection + SVM	Pearson Feature Selection + SVM	L1 Regularization Feature Selection + SVM	Random Forest Classification Feature Selection + SVM	Recursive Feature Elimination Feature Selection + SVM
5-fold cross-validation training	86.7%	92.7%	-	-	-	-	-
5-fold cross-validation test	70.1%	75.7%	-	-	-	-	-
The best model	C = 10, gamma = 0.083 (1/n_features * 10)	C = 1000, gamma = 0.083 (1/n_features * 10)	110 features C = 10, gamma = 0.09 (1/n_features * 10)	80 features C = 10, gamma = 0.125 (1/n_features * 10)	90 features C = 10, gamma = 0.11 (1/n_features * 10)	50 features C = 10, gamma = 0.2 (1/n_features * 10)	40 features C = 1000, gamma = 0.25 (1/n_features * 10)
Cross-validation grid search test	87.8%	90.3%	87.8%	89.1%	88.5%	87.7%	89.5%
F1 score	93%	93.5%	92.8%	92.8%	88.4%	92.8%	92.8%
ROC AUC score	92.7%	93.5%	92.1%	92.3%	87.7%	92.1%	92.7%
Accuracy	93%	93.5%	92.8%	92.8%	88.6%	92.8%	92.8%
Sensitivity	90%	93.5%	86%	88%	79.5%	85.5%	92%
Specificity	95.5%	93.5%	98.3%	96.7%	95.9%	98.7%	93.5%
Average F1 score (10 iterations)	91%	92.3%	-	-	-	-	-
Average ROC AUC score (10 iterations)	90.3%	92%	-	-	-	-	-
Average accuracy (10 iterations)	91%	92.4%	-	-	-	-	-
Average sensitivity (10 iterations)	83.9%	89%	-	-	-	-	-
Average specificity (10 iterations)	96.8%	95.1%	-	-	-	-	-

Table 3.12. The results for the SVM-based model - the case of the overlapping dataset [15]

	SVM	PCA Dimensionality Reduction + SVM	XGBoost Feature Selection + SVM	Pearson Feature Selection + SVM	L1 Regularization Feature Selection + SVM	Random Forest Classification Feature Selection + SVM	Recursive Feature Elimination Feature Selection + SVM
5-fold cross-validation training	73.6%	75.2%	-	-	-	-	-
5-fold cross-validation test	66.2%	67.2%	-	-	-	-	-
The best model	C = 10, gamma = 0.25 (1/n_features * 10)	C = 10, gamma = 0.25 (1/n_features * 10)	40 features C = 10, gamma = 0.25 (1/n_features * 10)	40 features C = 10, gamma = 0.25 (1/n_features * 10)	40 features C = 10, gamma = 0.25 (1/n_features * 10)	40 features C = 10, gamma = 0.25 (1/n_features * 10)	40 features C = 10, gamma = 0.25 (1/n_features * 10)
Cross-validation grid search test	85.3%	83.9%	85.1%	85.5%	85.5%	86.2%	86.1%
F1 score	88.9%	85.8%	89.2%	87.8%	88.1%	87%	87.5%
ROC AUC score	88.8%	86%	89.2%	87.8%	88%	86.9%	87.4%
Accuracy	88.9%	85.8%	89.2%	87.8%	88.1%	87%	87.5%
Sensitivity	87.4%	87.6%	89.2%	87.9%	86.7%	85.8%	86.8%
Specificity	90.2%	84.4%	89.1%	87.7%	89.3%	88%	88.1%
Average F1 score (10 iterations)	87.9%	86%	-	-	-	-	-
Average ROC AUC score (10 iterations)	87.8%	86.2%	-	-	-	-	-
Average accuracy (10 iterations)	87.9%	86%	-	-	-	-	-
Average sensitivity (10 iterations)	87%	88.6%	-	-	-	-	-
Average specificity (10 iterations)	88.6%	83.9%	-	-	-	-	-

For the non-overlapping dataset, we obtained the best performant SVM-based model (in the terms of the accuracy, 93.5%) with the configuration: SVM with penalty = 1000, and kernel coefficient = 1/(no_features*10), and PCA reduction. For the overlapping dataset, the configuration SVM with penalty =10 and kernel coefficient = 1/(no_features*10), and

XGBoost Feature selection generated the most performant model in terms of accuracy, 89.2%.

Table 3.13. The results for the DT models - the case of the non-overlapping dataset [15]

	GBT	RF
The best model	Iteration 9	max_depth = none max_features = log2 min_samples_split = 10 n_estimators = 300
Accuracy	**90.6%**	90.2%
Sensitivity	**84.5%**	83%
Specificity	**95.5%**	95.9%
F1 score	**90.5%**	90.1%
ROC AUC score	**90%**	89.4%

Table 3.14. The results for the DT models - the case of the overlapping dataset [15]

	GBT	RF
The best model	Iteration 7	max_depth = none max_features = sqrt min_samples_split = 10 n_estimators = 300
Accuracy	**92.2%**	90.3%
Sensitivity	**87.3%**	83.4%
Specificity	**96.2%**	95.8%
F1 score	**92.2%**	90.2%
ROC AUC score	**91.7%**	89.6%

The best performant DT-based models were obtained in the both cases, the overlapping and non-overlapping datasets using GBT.

Table 3.15. The results for the kNN model - the case of the non-overlapping dataset [15]

	kNN	PCA Dimensionality Reduction + kNN	XGBoost Feature Selection + kNN	Pearson Feature Selection + kNN	L1 Regularization Feature Selection + kNN	Random Forest Classification Feature Selection + kNN
5-fold cross-validation training	75.4%	73.3%	-	-	-	-
5-fold cross-validation test	61.5%	62.1%	-	-	-	-
The best model	leaf_size = 5, n_neighbors = 4, $p = 1$	leaf_size = 42, n_neighbors = 4, $p = 1$	30 features leaf_size = 5, n_neighbors = 3, $p = 1$	90 features leaf_size = 5, n_neighbors = 4, $p = 1$	80 features leaf_size = 5, n_neighbors = 4, $p = 1$	60 features leaf_size = 5, n_neighbors = 4, $p = 1$
Cross-validation grid search test	75.8%	72.8%	76.4%	75.7%	75.4%	76%
F1 score	80.2%	78.6%	80.3%	80.4%	79.5%	80.2%
ROC AUC score	80.8%	79.6%	81.4%	81%	80.1%	80.5%
Accuracy	80.1%	78.6%	80.4%	80.4%	79.5%	80.1%
Sensitivity	87%	89%	91%	87%	86%	83.5%
Specificity	74.6%	70.2%	71.8%	75.1%	74.2%	77.5%
Average F1 score (10 iterations)	77.2%	75.1%	-	-	-	-
Average ROC AUC score (10 iterations)	77.9%	76.6%	-	-	-	-
Average accuracy (10 iterations)	77.2%	75.3%	-	-	-	-
Average sensitivity (10 iterations)	84.4%	88.7%	-	-	-	-
Average specificity (10 iterations)	71.4%	64.6%	-	-	-	-

Table 3.16. The results for the kNN model - the case of the overlapping dataset [15]

	kNN	PCA Dimensionality Reduction + kNN	XGBoost Feature Selection + kNN	Pearson Feature Selection + kNN	L1 Regularization Feature Selection + kNN	Random Forest Classification Feature Selection + kNN
5-fold cross-validation training	81%	80.6%	-	-	-	-
5-fold cross-validation test	69.8%	69.3%	-	-	-	-
The best model	leaf_size = 5, n_neighbors = 4, $p = 1$	leaf_size = 6, n_neighbors = 4, $p = 1$	30 features leaf_size = 5, n_neighbors = 3, $p = 1$	40 features leaf_size = 5, n_neighbors = 3, $p = 1$	40 features leaf_size = 5, n_neighbors = 4, $p = 1$	30 features leaf_size = 5, n_neighbors = 3, $p = 1$
Cross-validation grid search test	78.6%	75.7%	79.9%	80%	79.8%	80.6%
F1 score	84%	78.7%	81.7%	80.2%	81%	81.8%
ROC AUC score	83.9%	78.7%	82.5%	81.3%	81%	82.5%
Accuracy	84%	78.7%	81.7%	80.2%	81%	81.7%
Sensitivity	83.3%	79.3%	90%	91.1%	81%	89.3%
Specificity	84.6%	78.3%	75.1%	71.4%	81%	75.7%
Average F1 score (10 iterations)	81.7%	77.8%	-	-	-	-
Average ROC AUC score (10 iterations)	81.8%	78.2%	-	-	-	-
Average accuracy (10 iterations)	81.7%	77.8%	-	-	-	-
Average sensitivity (10 iterations)	83.1%	81.1%	-	-	-	-
Average specificity (10 iterations)	80.6%	75.2%	-	-	-	-

In the case of kNN models, we obtained 84% accuracy for overlapping dataset and 80.4% accuracy using XGBoost or Pearson feature selection for non-overlapping datasets.

Table 3.17. The results for the NN models - the case of the non-overlapping dataset [15]

	Config. 1	Config. 2	Config. 3	Config. 4	Config. 5
F1 score	**87.5%**	85.7%	86.3%	85.2%	83.7%
ROC AUC score	**87.7%**	85.5%	86.1%	85.9%	83.6%
Accuracy	**87.5%**	85.7%	86.4%	85.3%	83.7%
Sensitivity	90.2%	83.4%	83.5%	**93.2%**	82.5%
Specificity	85.1%	87.6%	**89.7%**	78.6%	84.7%

For Configuration 3 (the case of the non-overlapping dataset), we applied five-fold cross validation (500 epochs with a batch size 32, Adam Optimizer with a decaying learning rate). The performance is shown in Table 3.18.

Table 3.18. Cross-validation and test score for Configuration 3 - the case of the non-overlapping dataset [15]

	Cross-validation	Test (10 runs averaged)
F1 score	97.4%	92.8%
ROC AUC score	97.5%	92.4%
Accuracy	97.4%	92.8%
Sensitivity	98.4%	87.8%
Specificity	96.6%	97.1%

Table 3.19. The results for the NN-based model - the case of the overlapping dataset [15]

	Config. 1	Config. 2	Config. 3	Config. 4	Config. 5	Config. 6	Config. 7
F1 score	81%	77.6%	80.2%	77.2%	75.7%	73.7%	67.2%
ROC AUC score	80.8%	77.4%	80.6%	77.6%	76%	75.6%	71.3%
Accuracy	81%	77.5%	80.1%	77.1%	75.6%	73.9%	68.5%
Sensitivity	79.3%	76.9%	84.5%	81.4%	79.2%	90%	94.9%
Specificity	82.4%	78%	76.7%	73.8%	72.7%	61.3%	47.8%

For the Configuration 2 (the case of the non-overlapping dataset), we applied five-fold cross-validation (200 epochs with a batch size 32, Adam Optimizer with a decaying learning rate). The model's performance is shown in Table 3.20.

Table 3.20. Cross-validation and test score for Configuration 2 - the case of the non-overlapping dataset [15]

	Cross-validation	Test (10 runs averaged)
F1 score	85.7%	78.3%
ROC AUC score	85.6%	78.8%
Accuracy	85.8%	78.2%
Sensitivity	83.8%	83.9%
Specificity	87.3%	73.7%

The algorithms provided statistically similar results. For both datasets, overlapping and non-overlapping datasets, the best performances – over 89% - were achieved by SVM and GBT algorithms. Considering ROC AUC score, we obtained the following best results:

- For the non-overlapping dataset: PCA reduction + SVM – 93.5%.
- For the overlapping dataset: GBT – 91.7%.

To provide the local explanations of the predictions we chose the LIME method, and we obtained the top 10 most relevant features (Figure 3.3-3.4).

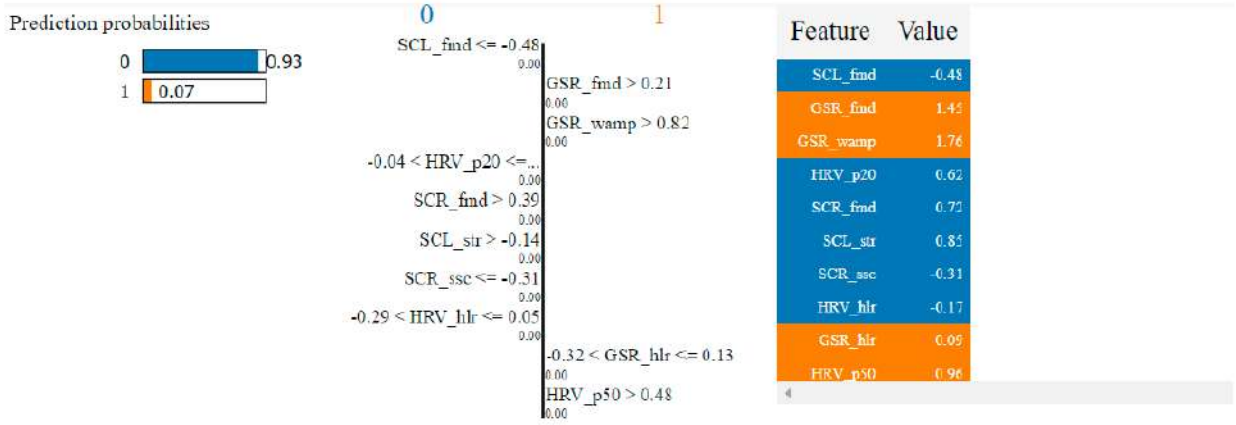

Figure 3.3. LIME interpretation for prediction non-fear for an instance (overlapping dataset) [15]

The features colored in blue contributed to prediction class 0.

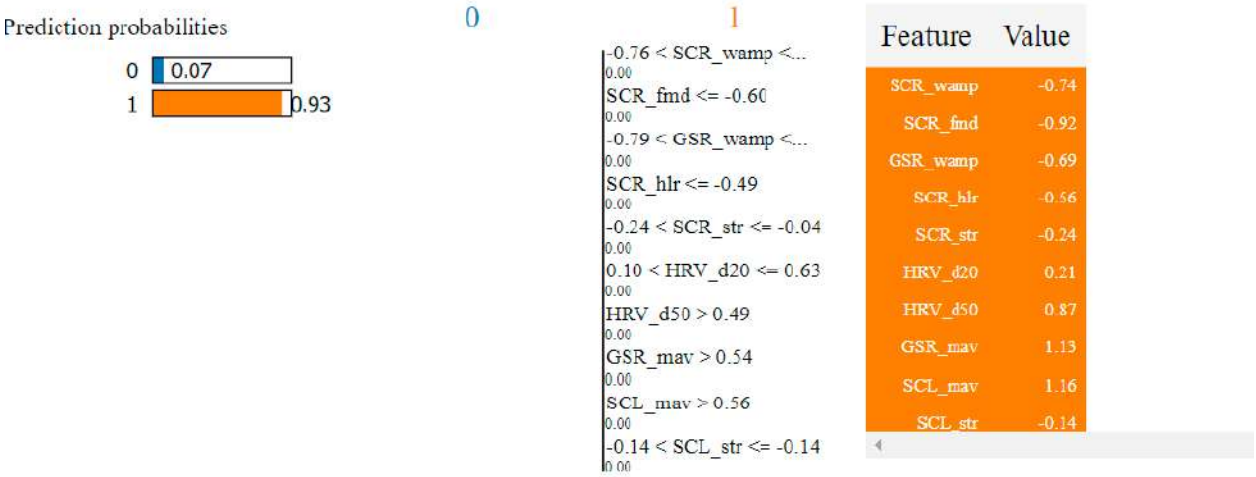

Figure 3.4. LIME interpretation for prediction fear for an instance (overlapping dataset) [15]

The features colored in orange contributed to prediction class 1.

The investigations performed on the DEAP dataset helped us further design our own experiments and build systems for treating fear of heights based on VRET systems.

Development of the adaptive VRET systems with emotion recognition

Our goal was to design and implement VR-based systems for acrophobia treatment, that adapt in real time to the emotional state of patients. We consider our systems to be AI-assistants, which help therapists in the phobia treatment process. In the ML-projects developments, we approached the strategy human-in-the-loop, in our case humans (therapists and patients) working together with AI to achieve more accurate models (through verifying the results, providing feedback, processing data).

To achieve our goal, we developed two VR-based games and gradually exposed subjects to various heights both in-vivo and in games. We performed more assisted experiments in the period 2018-2019 consisting in recording data from subjects, training and testing ML models for emotion recognition, and deployment and monitoring VRET systems. The main operations included in the systems were tracking the emotional states of the participants and automatically adjusting, in real time, the degree of exposure to the phobia triggers. In our case, for acrophobia treatment, we used the exposure to different levels of height in various VR-based landscape.

In [6], there are identified two main critical challenges regarding the building and deployment of the VRET systems with ML models: data used for training the ML models and the patients' safety. "The former regards the modality of emotion elicitation. Accordingly, we consider that the emotion-provoking stimuli should be more application-specific and they should be provided in the context of the experimental purpose. For our acrophobia application, the users should be exposed to realistic and immersive VR scenarios, have their biophysical data collected and supplied as training dataset. The latter is related to the patients' safe exposure, which means that the therapists should be engaged in the system's development" [6]. We addressed both challenges and acquired data in a specific context and the experiments were supervised by specialists.

To capture the EEG signals, we used an Acticap Xpress Bundle device with 16 dry electrodes positioned according the 10/20 system (https://trans-cranial.com/docs/10_20_pos_man_v1_0_pdf.pdf): FP1, FP2, FC5, FC1, FC2, FC6, T7, C3, C4, T8, P3, P1, P2, P4, O1 and O2 (Figure 3.5.). The letters show the brain area from which information is read: pre-frontal (Fp), frontal (F), temporal (T), parietal (P), occipital (O) and central (C). The even-numbered electrodes (2, 4, 6, 8) refer to electrode placement on the right side of the head, while odd numbers (1, 3, 5, 7) refer to those on the left. The ground and reference electrodes have been attached to the ears. The numbers 10 and 20 from the

10/20 standard signify that the distances between adjacent electrodes are either 10% or 20% of the total front-back or right-left distance of the skull. The GSR unit of a Shimmers Multi-Sensory device was used to record the EDA and HR.

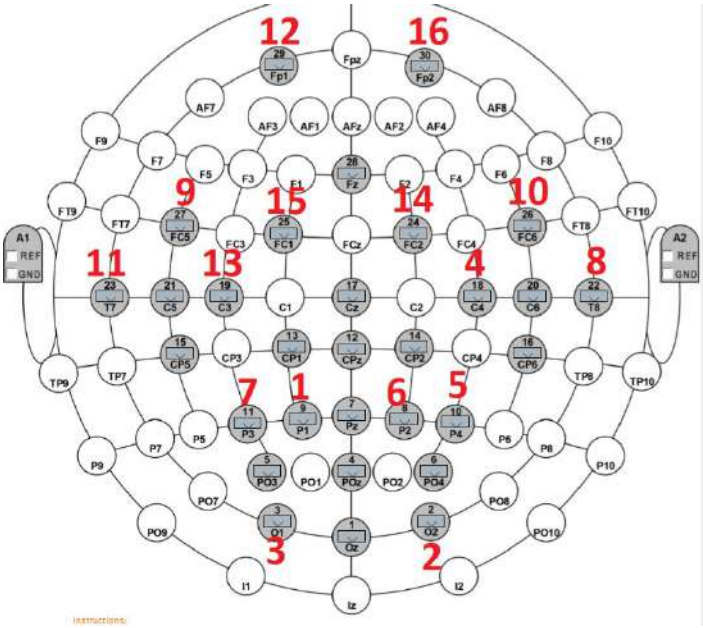

Figure 3.5. The electrodes' configuration to capture EEG signals (http://nets.cs.pub.ro/~safe_vr/experiment.html)

The protocols for measuring data and conducting experiments were modified depending on the software used. In the first studies, EEG, GSR, and HR data were recorded. Over time, EEG data recording was discontinued due to the uncomfortable measuring device and respiratory rate measurement was introduced.

ML models for fear recognition in the VRET systems

Our approach for integrating AER in VRET for acrophobia treatment consisted in developing machine learning models for fear recognition and game level prediction and embedded them in VR based video games [5], [6], [9], [10].

In the first VR-based game [5], [6], [9], the players are exposed in real-time to appropriate in-game height levels according to physiological data acquired from the players. Two Deep Neural Networks (DNNs) were built and integrated in the system, one to estimate the current fear level of the players and one to predict the next game level that is displayed to the players.

The architecture of the proposed VRET system is presented in Figure 3.6. A module of data acquisition (EEG, HR and GSR signals) collects, pre-processes and transfers data to the DBMS (Database Management System). The user plays the video game and self-assesses their level of fear at the current level of play, called Subjective Unit of Distress (SUD). Data extracted from EEG, HR and GSR signals feeds a deep neural network (DNN1) to determine the current level of fear. SUD is used to validate the accuracy for DNN1. Using EEG, biophysical data and the desired level of fear, DNN2 outputs the next game level suitable for a specific user.

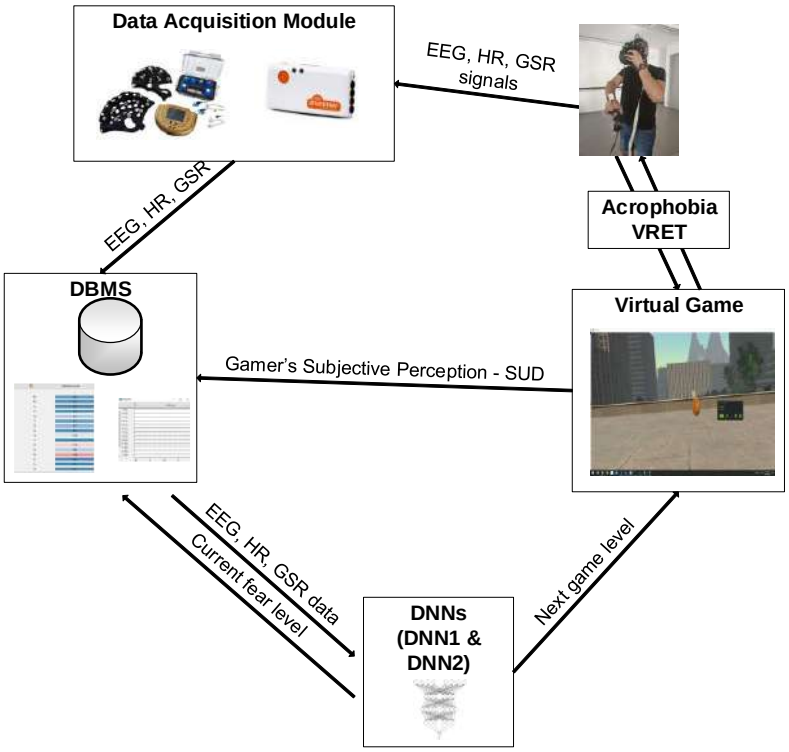

Figure 3.6. Architecture of the proposed emotion-aware VRET system [9]

We use the following notations:

n – numbers of levels of the game

$l_0, l_1, \dots, l_{n-1}$ – ordered game levels corresponding to a degree of height exposure

l_{cr} – the current game level

l_{pr} – the predicted game level

fl_{cr} – the current fear level

fl_d – the next desired fear level

The work scenario is presented in Figure 3.7. The user starts playing l_0 level, so the current level of game l_{cr} is l_0. The EEG, HR and GSR data are recorded during the game. When the user finishes a game level, the acquired data fed DNN1 and the current level of fear is

predicted. The next desired level of fear is computed according to the 2-choices scale or 4-choices scale paradigms.

In the case of 2-choices scale, we use the following strategy to determine the desired fear level:

if $fl_{cr} == 0$ then $fl_d = 0$

if $fl_{cr} == 1$ then $fl_d = 1$

In the case of 4-choices scale, we use the following strategy to determine the desired fear level:

if $fl_{cr} == 0$ or $fl_{cr} == 1$ then $fl_d = fl_{cr} + 1$

if $fl_{cr} == 2$ then $fl_d = fl_{cr}$

if $fl_{cr} == 3$ then $fl_d = fl_{cr} - 1$

The desired fear level and biophysical data are inputs to the DNN2, and a game level is predicted (l_{pr}). The users play the predicted level of the game and once again their biophysical data are recorded. DNN1 outputs the general fear level and DNN2 predicts the game level. The algorithm stops when a predefined number of epochs is reached.

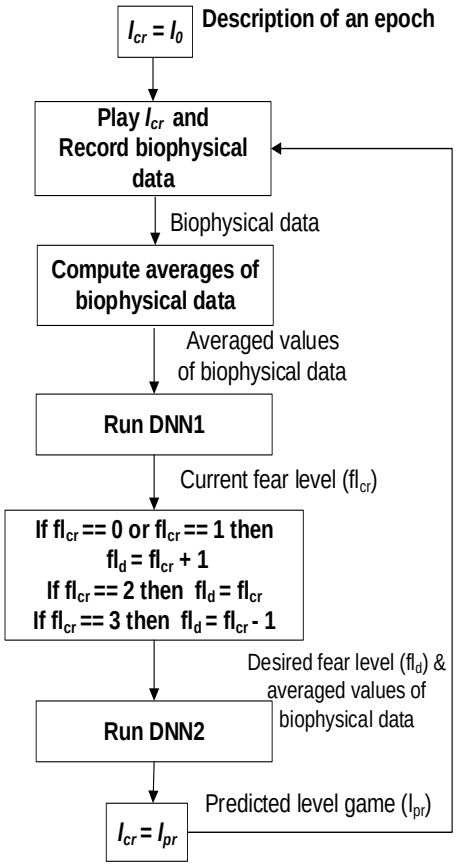

Figure 3.7. The algorithm in the case of 4-choices scale [9]

The video game shows an urban landscape with tall buildings and streets. Four subjects (acrophobic patients) played the game twice, consisting in exploring various floors of the

building and collecting coins situated at different distances from the edge of the balconies. A collage from the video game is presented in Figure 3.8.

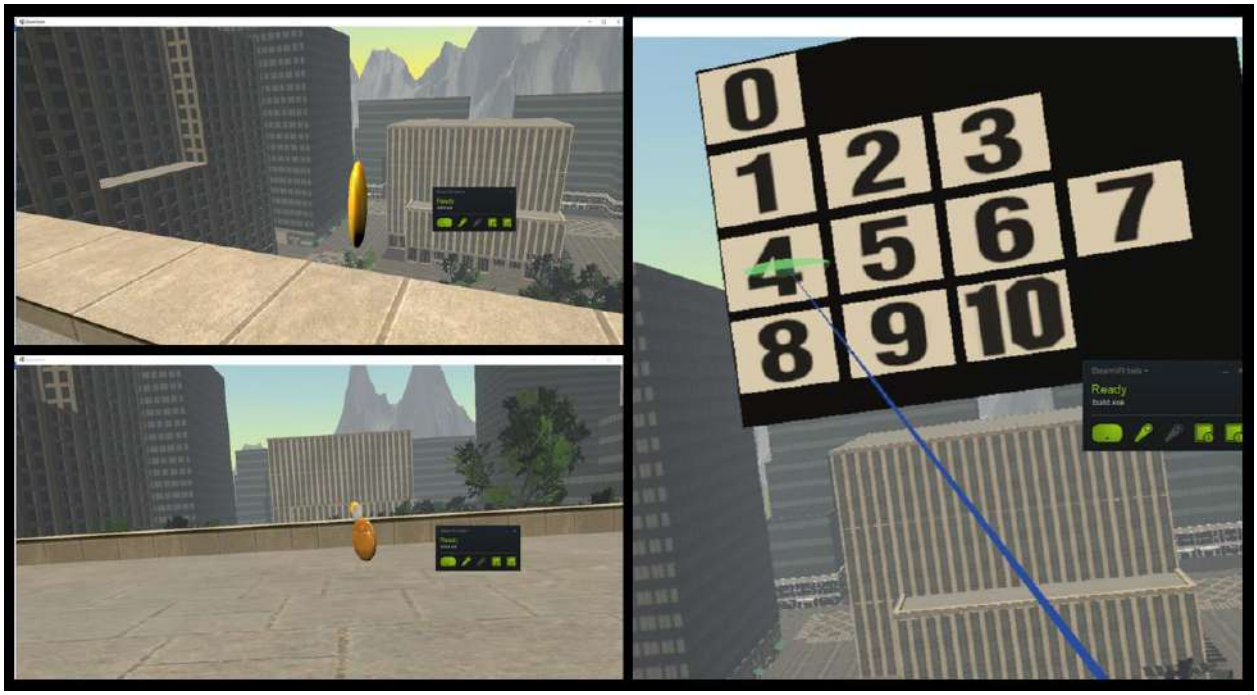

Figure 3.8. A collage from the video game (http://nets.cs.pub.ro/~safe_vr/game.html)

We collected data in both virtual and real-conditions and acquired 63 trials per subject. As such, our dataset contains 25000 entries on average for each subject, recorded at intervals of 65 ms. The alpha, beta and theta log-normalized powers for all channels, as well as the ratio of the theta to the beta powers (slow waves/fast waves) were computed.

Two DNN models were used:

- DNN1 for emotion recognition (current fear level).
- DNN2 to predict the next level of the game.

To obtain datasets, we performed more preliminary experiments exposing the subjects to various heights, both in-vivo and virtual experiments. During recording data, the user reported SUD on 11-choices scale (0 – for complete relaxation and 10 for extreme emotion and panic attack). In the real-world situations, a baseline was established by realising measurements where the patient is at complete relaxation, at the first, fourth, and sixth floors of a building and at about 4m, 2m, and a few centimetres away from the balcony edge. There were two exposures in-vivo for each subject, before and after virtual exposure.

Because the subjects assessed their level of fear from 0 to 10, we translated the 11-choices scale to 2-choices scale and 4-choices scale according to the following technique (Table 3.21).

Table 3.21. The translating of the 11-choices scale into 2-choice scale and 4-choices scale

11-choices scale	0	1	2	3	4	5	6	7	8	9	10
4-choices scale	0	1			2				3		
2-choices scale	0				1						

We consider the following meaning for the 2-choices scale and 4-choices scale:

- 2-choices scale, 0 –relaxation and 1 - fear.
- 4-choices scale, 0 - for complete relaxation, 1 - low fear, 2 - moderate fear and 3 - high level of anxiety.

DNN1 – for emotion recognition

In total we have 71 input features:

- Alpha, beta, theta, theta/beta values from channels: FP1, FP2, FC5, FC1, FC2, FC6, T7, C3, C4, T8, P3, P1, P2, P4, O1, O2.
- Values for: Alpha FP2- Alpha FP1; (Alpha FC2 + Alpha FC6) – (Alpha FC5 +Alpha FC1); (Alpha FP2 + Alpha FC2 + Alpha FC6) – (Alpha FP1 + Alpha FC5 + Alpha FC1); (Beta FP2 + Beta FC2 + Beta FC6) – (baseline Beta FP2 + baseline Beta FC2 + baseline Beta FC6); (theta/beta FC6+ theta/beta FC2 + theta/beta FP2) – (baseline theta/beta FC6+ baseline theta/beta FC2 + baseline theta/beta FP2).
- GSR and HR values.

The targets are the values for fear levels on 11-choices scale (with values from 0 to 10). 4-choices scale (values from 0 to 3), 2-choices scale (values of 0 or 1).

To obtain the best model for our AER, we developed four MLP sequential models for binary and multi-class classification (Table 3.22.).

Table 3.22. AER models for the VRET system [9]

	Model 1	Model 2	Model 3	Model 4
	3 hidden layers – 150 neurons on each layer, RELU activation	3 hidden layers – 300 neurons on each layer, RELU activation	6 hidden layers – 150 neurons on each layer, RELU activation	6 hidden layers – 300 neurons on each layer RELU activation
2 - choices scale	Output layer – Sigmoid activation Binary cross-	Output layer – Sigmoid activation Binary cross-	Output layer – Sigmoid activation Binary cross	Output layer – Sigmoid activation Binary cross-

	entropy loss function	entropy loss function	entropy loss function	entropy loss function
4 – choices scale	Output layer - Softmax activation function Logarithmical categorical cross-entropy loss function	Output layer - Softmax activation function Logarithmical categorical cross-entropy loss function	Output layer - Softmax activation function Logarithmical categorical cross-entropy loss function	Output layer - Softmax activation function Logarithmical categorical cross-entropy loss function
11 – choices scale	Output layer - Softmax activation function Logarithmical categorical cross-entropy loss function	Output layer - Softmax activation function Logarithmical categorical cross-entropy loss function	Output layer - Softmax activation function Logarithmical categorical cross-entropy loss function	Output layer - Softmax activation function Logarithmical categorical cross-entropy loss function

For all models, we used Adam gradient descendent optimization algorithm, 1000 epochs for training and a batch size of 20. The performances for the AER models are presented in Table 3.23.

Table 3.23. Performances of AER models [9]

DNN1			
Model	Maximum cross-validation accuracy (%)		
	2-choices scale	4-choices scale	11-choices scale
Model_1	95.03	87.94	**85.09**
Model_2	**95.51**	**90.49**	79.48
Model_3	94.43	86.32	74.27
Model_4	94.57	88.28	80.45

In the Tables 3.24. - 3.25., we present some subject's data recorded in different situations.

Table 3.24. Example of data - baseline

ALPHA															
FP1	FP2	FC5	FC1	FC2	FC6	T7	C3	C4	T8	P3	P1	P2	P4	O1	O2
4.468987	4.638647	2.298466	2.467224	4.46244	3.726878	3.624476	2.283079	3.726878	2.138125	2.149194	3.541537	3.726878	4.250161	3.726878	3.268173

BETA															
FP1	FP2	FC5	FC1	FC2	FC6	T7	C3	C4	T8	P3	P1	P2	P4	O1	O2
3.17242	3.252612	3.567698	3.775043	5.389293	5.382977	5.005379	3.743094	5.382977	3.714502	4.438111	4.094626	5.382977	5.373084	5.382977	3.815996

THETA															
FP1	FP2	FC5	FC1	FC2	FC6	T7	C3	C4	T8	P3	P1	P2	P4	O1	O2
4.243558	4.768674	2.784969	3.061673	4.832994	3.745596	3.693426	2.899508	3.745596	2.955975	2.690639	5.13273	3.745596	4.4288	3.745596	4.75521

THETA/BETA															
FP1	FP2	FC5	FC1	FC2	FC6	T7	C3	C4	T8	P3	P1	P2	P4	O1	O2
0.66882	0.733053	0.780607	0.81103	0.896777	0.695822	0.737891	0.774629	0.695822	0.795793	0.606258	1.253528	0.695822	0.824257	0.695822	1.246125

GSR	Pulse
0.99517516	75.24750149

Table 3.25. Example of data - various level of heights in VR conditions

	ALPHA															
Floor / position	FP1	FP2	FC5	FC1	FC2	FC6	T7	C3	C4	T8	P3	P1	P2	P4	O1	O2
P--1	3.503892	2.896083	2.495974	2.275845	2.3931	2.550224	3.535742	2.070293	3.987501	2.083388	3.13693	2.527608	3.987501	2.540366	3.987501	3.090283
P--2	3.239585	2.996605	2.785175	2.22867	2.62501	2.70819	3.284469	2.115948	3.980763	2.138393	2.241442	2.10786	3.980763	2.363486	3.980763	2.50996
P--3	3.301529	2.662959	2.754192	2.274489	2.252268	2.50137	3.276848	2.118844	3.973349	2.152324	2.214148	1.976768	3.973349	2.487142	3.973349	2.42714
1--1	3.212697	2.863069	5.706423	2.115041	2.031369	2.226406	3.348228	1.762564	3.965235	1.695213	1.795213	2.037601	3.965235	2.338431	3.965235	2.331008
1--2	3.40892	2.893017	6.686761	2.594526	2.72538	2.750569	3.364398	2.185702	3.956601	2.203633	2.240396	1.935961	3.956601	2.612176	3.956601	2.346245
1--3	3.208121	2.622227	5.370759	2.388889	2.315563	2.457556	3.204464	2.00252	3.947659	2.193241	2.297086	2.288211	3.947659	2.574624	3.947659	2.511906
4--1	2.935946	2.509878	3.686198	2.388428	1.999771	2.067191	3.099483	1.699334	3.938778	1.902968	2.060691	2.758177	3.938778	2.360999	3.938778	2.729347
4--2	2.966746	2.107847	2.878077	1.822348	2.031287	2.113311	2.611677	1.969957	3.930074	1.659816	2.032511	2.482623	3.930074	2.147753	3.930074	2.597208
4--3	3.058292	2.727784	2.850501	2.197695	2.176509	2.261191	3.02155	1.867368	3.92135	1.953932	1.903344	2.904563	3.92135	2.343217	3.92135	2.207048
6--1	3.12524	2.558639	2.681191	2.036931	2.11049	2.343229	3.182617	1.764988	3.912493	1.892438	1.895786	2.239401	3.912493	2.171497	3.912493	2.29269
6--2	3.028642	2.536119	2.64066	2.123024	2.134253	2.103594	3.030362	2.025526	3.903561	2.242655	2.099304	2.56625	3.903561	2.353571	3.903561	2.669227
6--3	3.337986	3.111522	2.841108	2.123137	2.300009	2.375259	3.267446	1.914309	3.894857	2.0461	2.295186	2.284584	3.894857	2.32404	3.894857	2.334306
10--1	2.992564	2.794926	2.301249	2.197942	2.294971	2.368559	2.933374	2.013791	3.886676	2.118887	1.744859	3.038971	3.886676	2.446347	3.886676	3.146819
10--2	3.170476	2.709144	2.272261	2.203872	2.262553	2.366289	3.362912	2.046273	3.879698	2.147321	2.019485	2.438252	3.879698	2.446323	3.879698	2.61704
10--3	2.640202	2.695827	2.704718	2.097439	2.358461	2.299832	3.17162	2.004463	3.873938	1.82279	1.90073	2.221822	3.873938	2.262991	3.873938	2.424596

	BETA															
Floor / position	FP1	FP2	FC5	FC1	FC2	FC6	T7	C3	C4	T8	P3	P1	P2	P4	O1	O2
P--1	4.618633	3.47275	3.55183	3.239347	3.3192	3.375258	4.758495	3.299274	4.851112	3.309229	5.417393	4.022051	4.851112	3.391792	4.851112	4.097561
P--2	4.606552	3.239785	3.630352	3.267257	3.302454	3.420189	4.727211	3.149247	4.848784	3.236227	4.643969	3.866982	4.848784	3.207377	4.848784	4.367567
P--3	4.450975	3.142763	5.701455	3.123225	3.025884	3.269659	4.646999	3.016983	4.846335	2.975916	4.554029	3.714423	4.846335	3.114384	4.846335	3.582095
1--1	4.358903	3.189744	7.89474	3.206424	3.044845	3.200842	4.589822	3.003222	4.843907	3.05129	4.569865	3.826121	4.843907	3.126089	4.843907	3.363114
1--2	4.457994	3.286521	6.082603	3.16214	3.056854	3.256254	4.578893	2.834355	4.841438	3.058128	4.407845	3.793256	4.841438	3.124873	4.841438	3.181035
1--3	4.50599	3.274726	8.333407	3.822051	3.125692	3.31395	4.694619	2.998765	4.838853	2.986162	4.379901	3.867377	4.838853	2.955636	4.838853	3.117164
4--1	4.416806	3.252652	5.708298	3.342881	2.986851	3.048102	4.44233	3.00606	4.836299	3.062008	4.435668	3.997478	4.836299	3.172729	4.836299	3.296736
4--2	4.389148	3.166237	3.840948	3.382591	3.098846	3.194131	4.514276	3.085047	4.83377	3.016904	4.718101	3.986087	4.83377	3.023555	4.83377	3.133721
4--3	4.358817	3.299783	3.813453	3.068126	3.05364	3.15128	4.381888	3.078379	4.831428	3.055119	4.537778	4.006757	4.831428	3.08174	4.831428	3.243721
6--1	4.423331	3.27733	3.462981	3.013182	3.062171	3.157766	4.32363	3.118601	4.829212	3.155631	4.677478	3.976257	4.829212	3.068467	4.829212	3.306636
6--2	4.297079	3.138967	3.112862	3.194021	3.141208	3.171565	4.486491	2.895047	4.827175	3.080029	4.588141	3.774486	4.827175	3.205023	4.827175	3.172039
6--3	4.418303	3.350956	3.32565	3.039699	3.094268	3.188412	4.447568	3.058123	4.825415	3.110848	4.676378	3.99696	4.825415	3.068712	4.825415	3.14592
10--1	4.3855	3.385479	3.345734	3.228794	3.20689	3.227992	4.401239	3.14084	4.823459	3.128758	4.509255	4.035366	4.823459	3.21246	4.823459	3.453393
10--2	4.242225	3.29034	3.337974	3.115974	3.123498	3.229631	4.626202	3.087579	4.821583	3.082189	4.561973	4.006128	4.821583	3.215274	4.821583	3.393697
10--3	4.178984	3.275122	3.290012	2.940296	3.766484	3.11169	4.519214	3.07071	4.819778	3.101287	4.661328	3.960938	4.819778	2.995211	4.819778	3.046756

	THETA															
Floor / position	FP1	FP2	FC5	FC1	FC2	FC6	T7	C3	C4	T8	P3	P1	P2	P4	O1	O2
P--1	3.538713	3.33144	2.688128	2.818	2.692066	2.588214	3.47043	2.739468	4.128596	2.775865	3.141565	3.297816	4.128596	3.04521	4.128596	3.717094
P--2	3.485341	3.372254	2.835505	2.355925	2.446638	2.386481	3.088245	2.182602	4.123871	2.202288	2.217903	2.568155	4.123871	2.477055	4.123871	2.753987
P--3	3.344729	3.461774	3.072519	2.724433	2.706508	2.618296	3.148236	2.372184	4.118853	2.527755	2.32331	2.49283	4.118853	2.783571	4.118853	2.942718
1--1	3.393134	3.177823	5.297091	2.58177	2.455458	2.614162	2.953454	2.059013	4.113882	2.381683	2.085021	2.995006	4.113882	2.6154	4.113882	3.032508
1--2	3.751017	3.402998	7.728143	2.818082	2.767495	2.781648	3.293464	2.308754	4.108868	2.506172	2.32915	2.604572	4.108868	3.079552	4.108868	2.889817
1--3	3.307828	3.054032	4.444277	2.596678	2.69652	2.673747	3.240752	2.178584	4.103751	2.416552	2.095242	2.61343	4.103751	2.703424	4.103751	2.848416
4--1	3.42708	3.156117	4.253226	3.29464	2.872419	2.876512	3.239877	2.594525	4.097663	2.732549	2.194321	3.925198	4.097663	3.175717	4.097663	3.800529
4--2	3.368232	3.164815	3.148966	2.786754	2.70003	2.653401	3.227547	2.536134	4.091051	2.593576	2.28082	3.665514	4.091051	2.862174	4.091051	3.461825
4--3	3.15344	3.534842	2.83946	2.919997	2.90997	2.926642	3.125613	2.488945	4.08344	2.651107	2.199091	4.74313	4.08344	3.082347	4.08344	2.867921
6--1	3.459145	2.974713	2.891284	2.552031	2.480189	2.611378	3.178245	2.1538	4.075978	2.455221	2.044302	2.916431	4.075978	2.853815	4.075978	3.055786
6--2	3.368109	3.503512	2.76068	2.537743	2.638999	2.868467	3.092306	2.32921	4.068963	2.58895	2.238169	3.416006	4.068963	3.044016	4.068963	3.517006
6--3	3.405319	3.325513	2.580075	2.639845	2.844208	2.737339	3.138856	2.321648	4.062002	2.477976	2.20444	2.439823	4.062002	2.750602	4.062002	2.759415
10--1	3.186336	3.137052	2.687873	2.668974	2.573863	2.618858	3.154834	2.217302	4.055249	2.74627	2.015562	4.370573	4.055249	3.125593	4.055249	4.419974
10--2	3.34206	3.336075	2.686366	2.780297	2.831684	2.620966	3.470362	2.452675	4.048823	2.703454	1.877235	3.398713	4.048823	2.877075	4.048823	3.444012
10--3	3.150935	3.235537	2.805456	2.475659	2.598132	2.481385	3.129612	2.210556	4.042164	2.467428	1.928989	3.160882	4.042164	2.743378	4.042164	3.090173

Floor / position	THETA/BETA															
	FP1	FP2	FC5	FC1	FC2	FC6	T7	C3	C4	T8	P3	P1	P2	P4	O1	O2
P--1	0.766182	0.959309	0.756829	0.869928	0.811059	0.766819	0.729313	0.830324	0.851062	0.838825	0.579903	0.819934	0.851062	0.897817	0.851062	0.907148
P--2	0.756605	1.040888	0.781055	0.721071	0.740854	0.697763	0.653291	0.693055	0.850496	0.680511	0.477588	0.664124	0.850496	0.772299	0.850496	0.630554
P--3	0.75146	1.101507	0.538901	0.872314	0.894452	0.800786	0.677477	0.786277	0.84989	0.849404	0.510166	0.671122	0.84989	0.893779	0.84989	0.821507
1--1	0.778437	0.996262	0.670965	0.805187	0.806431	0.816711	0.643479	0.685601	0.84929	0.78055	0.456254	0.782779	0.84929	0.836636	0.84929	0.901696
1--2	0.841414	1.035441	1.270532	0.891195	0.905341	0.854248	0.719271	0.814561	0.848687	0.819512	0.52841	0.686632	0.848687	0.985497	0.848687	0.908452
1--3	0.734096	0.932607	0.533309	0.679394	0.862695	0.806816	0.690312	0.726494	0.848083	0.80925	0.478376	0.675763	0.848083	0.914667	0.848083	0.913784
4--1	0.775918	0.970321	0.745095	0.985569	0.961688	0.943706	0.729319	0.863098	0.847272	0.892404	0.494699	0.981919	0.847272	1.000942	0.847272	1.152816
4--2	0.7674	0.999551	0.819841	0.823852	0.871302	0.830711	0.714965	0.822073	0.846348	0.859681	0.483419	0.919577	0.846348	0.946625	0.846348	1.104701
4--3	0.723462	1.071235	0.74459	0.95172	0.952951	0.928715	0.713303	0.808525	0.845183	0.867759	0.484619	1.183783	0.845183	1.000197	0.845183	0.884145
6--1	0.782023	0.907663	0.834912	0.846955	0.809945	0.82697	0.735087	0.69063	0.844025	0.778044	0.437052	0.733461	0.844025	0.930046	0.844025	0.924138
6--2	0.783814	1.116135	0.886862	0.794529	0.840122	0.904433	0.689248	0.80455	0.842928	0.84056	0.487816	0.905026	0.842928	0.949764	0.842928	1.108753
6--3	0.77073	0.992407	0.775811	0.868456	0.919186	0.858527	0.705747	0.759174	0.841793	0.79656	0.471399	0.61042	0.841793	0.896338	0.841793	0.877141
10--1	0.726562	0.92662	0.803373	0.826616	0.802604	0.811296	0.716806	0.705958	0.840735	0.877751	0.446983	1.083067	0.840735	0.972959	0.840735	1.279893
10--2	0.787808	1.013899	0.804789	0.892272	0.906575	0.811537	0.750154	0.794368	0.839729	0.877122	0.411496	0.848379	0.839729	0.894815	0.839729	1.014826
10--3	0.753995	0.987913	0.852719	0.841976	0.689803	0.797439	0.692512	0.719885	0.838662	0.795614	0.413828	0.798014	0.838662	0.915921	0.838662	1.01425

Floor / position	Alpha FPR-FPL	Alpha FCR-FCL	Alpha FR-FL	Beta diff from baseline FR	Theta/beta diff from baseline FR	GSR	Pulse
P--1	-0.607808661	0.171504407	-0.4363043	-3.857673758	0.211534799	1.486159	73.36841
P--2	-0.242979679	0.319355303	0.0763756	-4.062454022	0.153853358	1.444329	67.67613
P--3	-0.638570247	-0.275043062	-0.9136133	-4.586576597	0.471092394	1.418664	66.45312
1--1	-0.349627674	-3.563689651	-3.9133173	-4.589450323	0.293752013	1.390636	69.02582
1--2	-0.515903769	-3.805337784	-4.3212416	-4.425252554	0.469377331	1.364219	68.28457
1--3	-0.585894078	-2.986528182	-3.5724223	-4.310514232	0.276465667	1.340664	72.61482
4--1	-0.426068191	-2.007663881	-2.4337321	-4.737276797	0.550062805	1.308326	71.89398
4--2	-0.858898555	-0.555826627	-1.4147252	-4.565666875	0.375911409	1.294569	74.20773
4--3	-0.330507571	-0.610494823	-0.9410024	-4.520178906	0.62724903	1.27974	72.38466
6--1	-0.566601646	-0.264402956	-0.8310046	-4.527615106	0.218926114	1.356619	74.66977
6--2	-0.49252286	-0.525837582	-1.0183604	-4.573141909	0.535038258	1.363816	76.31953
6--3	-0.226464463	-0.288976706	-0.5154412	-4.391246415	0.444468682	1.33841	73.83352
10--1	-0.19763797	0.164338324	-0.0332996	-4.204520328	0.214867578	1.327102	76.20389
10--2	-0.461332284	0.152709188	-0.3086231	-4.381411856	0.406358967	1.308009	75.22607
10--3	0.055624297	-0.143864266	-0.08824	-3.871585088	0.149503552	1.366783	77.34429

Alpha FPR-FPL	Alpha FCR-FCL	Alpha FR-FL	Beta diff from baseline FR	Theta/beta diff from baseline FR	GSR	Pulse	RESPONSE (L11)	RESPONSE (0/1) - L2	RESPONSE (0/1/2/3) - L4
-0.607808661	0.171504407	-0.4363043	-3.857673758	0.211534799	1.486158798	73.3684076	0	0	0
-0.242979679	0.319355303	0.0763756	-4.062454022	0.153853358	1.444328739	67.67612797	0	0	0
-0.638570247	-0.275043062	-0.9136133	-4.586576597	0.471092394	1.418663578	66.45312054	0	0	0
-0.349627674	-3.563689651	-3.9133173	-4.589450323	0.293752013	1.390635922	69.02582047	0	0	0
-0.515903769	-3.805337784	-4.3212416	-4.425252554	0.469377331	1.364219193	68.28456691	0	0	0
-0.585894078	-2.986528182	-3.5724223	-4.310514232	0.276465667	1.3406644	72.61481918	1	0	1
-0.426068191	-2.007663881	-2.4337321	-4.737276797	0.550062805	1.308325848	71.89397946	0	0	0
-0.858898555	-0.555826627	-1.4147252	-4.565666875	0.375911409	1.294569312	74.20772618	1	0	1
-0.330507571	-0.610494823	-0.9410024	-4.520178906	0.62724903	1.279740487	72.38466197	2	0	1
-0.566601646	-0.264402956	-0.8310046	-4.527615106	0.218926114	1.356619117	74.66977011	1	0	1
-0.49252286	-0.525837582	-1.0183604	-4.573141909	0.535038258	1.363815672	76.31952988	2	0	1
-0.226464463	-0.288976706	-0.5154412	-4.391246415	0.444468682	1.338409837	73.83352097	3	0	1
-0.19763797	0.164338324	-0.0332996	-4.204520328	0.214867578	1.327101614	76.20389466	3	0	1
-0.461332284	0.152709188	-0.3086231	-4.381411856	0.406358967	1.308009157	75.22606567	4	1	2
0.055624297	-0.143864266	-0.08824	-3.871585088	0.149503552	1.366783397	77.34428865	4	1	2
-0.713337686	-2.506458632	-3.2197963	-2.229967207	0.555115551	0.24100294	72.08857204	0	0	0
-1.032133996	-2.319719798	-3.3518538	-2.528637913	0.460422927	0.239668893	75.75103821	0	0	0
-0.885958509	-2.28047503	-3.1664335	-2.828121692	0.310191437	0.238641403	68.94700561	0	0	0

DNN2 - models for game levels prediction

Inputs: EEG, GSR, HR and SUD values

Output: height level – 0 for ground floor, 1 for the first floor, 2 for the fourth floor, 3 for the sixth floor and 4 for the eighth floor.

To obtain the best model for the DNN2, we developed four MLP sequential models (Table 3.26).

Table 3.26. DNN models for game level prediction [9]

Model 1	Model 2	Model 3	Model 4
3 hidden layers – 150 neurons on each layer, RELU activation Output layer with 5 neurons, Softmax activation function Logarithmical categorical cross-entropy loss function	3 hidden layers – 300 neurons on each layer, RELU activation Output layer with 5 neurons, Softmax activation function Logarithmical categorical cross-entropy loss function	6 hidden layers – 150 neurons on each layer, RELU activation Output layer with 5 neurons, Softmax activation function Logarithmical categorical cross-entropy loss function	6 hidden layers – 300 neurons on each layer RELU activation Output layer with 5 neurons, Softmax activation function Logarithmical categorical cross-entropy loss function

The performances for the ML models for game level prediction are presented in Table 3.27.

Table 3.27. Performances for ML models for game level prediction [9]

DNN2			
Model	Maximum cross-validation accuracy (%)		
	2-choices scale	4-choices scale	11-choices scale
Model_1	98.40	**98.67**	**98.75**
Model_2	**98.72**	98.50	98.65
Model_3	97.45	97.82	98.50
Model_4	97.37	97.77	98.17

Some captures from the training data sets are presented in Table 3.28.

Table 3.28. Captures from the datasets used for DNN2 models

FP1	FP2	FC5	FC1	FC2	FC6	T7	C3	C4	T8	P3	P1	P2	P4	O1	O2
5.867373	4.227321	4.035462	4.251962	2.96355	4.889579	3.749226	2.690703	5.470478	2.335237	3.458314	3.104926	3.056246	2.805908	5.470478	2.969957
6.009617	3.852959	4.152323	3.770973	3.072012	4.752671	3.794914	3.265575	6.402118	2.53816	3.663238	3.379235	3.108266	2.924835	6.402118	3.072226
7.04568	4.767462	4.493221	3.696241	3.846509	5.305697	4.23678	3.282405	5.662794	2.737701	3.792534	3.19194	3.362033	3.648162	5.662794	3.029228
5.425718	5.984514	6.400675	4.838521	5.498883	5.477169	5.191879	4.510276	8.930783	4.772232	4.423476	4.85785	8.930783	4.943974	8.930783	5.006512
5.264197	5.818885	6.124851	4.209558	5.49546	5.542967	5.107944	4.375087	6.984232	4.408911	4.412819	4.128713	6.984232	4.516275	6.984232	4.195445
6.384926	6.904209	7.187443	4.992828	5.879195	5.946026	5.639377	4.977989	7.560023	4.727118	4.544573	4.729427	7.560023	4.882753	7.560023	4.579535
6.025629	6.200857	5.5398	4.97467	4.623609	7.630639	4.354375	3.064053	5.04728	3.199221	2.877751	2.918963	3.029128	4.208601	5.04728	3.587385
6.637346	6.524375	5.959029	5.564565	5.157339	8.506503	4.847374	3.183234	5.655232	3.317747	3.10725	2.938554	3.400843	4.441916	5.655232	3.673813
6.970063	7.131989	6.269422	6.190973	5.497449	9.019936	5.961138	3.284148	8.22335	3.533857	3.911776	3.145804	3.571473	4.355067	8.22335	3.88144
3.575436	3.319158	3.470367	2.713542	4.571692	7.749039	3.793566	2.379283	6.9357	3.343694	3.732986	3.020053	6.9357	2.944842	6.9357	2.802728
3.077304	3.377111	3.581184	2.499254	4.682971	6.789535	3.242271	2.04768	5.789918	2.953746	3.199195	2.690203	5.789918	2.509736	5.789918	2.494359
3.747682	4.741622	4.67252	3.215557	5.64401	7.887133	3.570875	2.037467	12.13106	3.138179	3.605912	3.090809	12.13106	3.462038	12.13106	3.024651
3.240678	4.211957	4.162693	2.939405	4.72923	7.673064	3.244547	2.188977	8.084619	3.111237	3.443653	3.067501	8.084619	2.931274	8.084619	2.568779
3.593475	4.437443	4.463479	3.250363	4.894451	8.557459	3.423814	3.544095	7.376502	3.081403	3.849519	3.467501	7.376502	3.523858	7.376502	2.793589
3.736109	3.533609	3.856748	2.838578	4.108795	9.180539	2.982244	2.211896	4.854611	2.920364	3.094098	3.05153	4.854611	2.946485	4.854611	2.910147
4.025368	3.475151	3.341852	2.722136	4.101204	7.0766	3.089884	2.505091	5.138336	3.308655	3.260412	3.072865	5.138336	3.123146	5.138336	2.971915
4.069512	4.102048	4.540406	3.251883	4.879988	7.841659	3.611733	2.637259	5.590758	3.296236	3.749141	3.677058	5.590758	4.214297	5.590758	3.418819
3.646828	3.674215	3.869561	3.354286	4.176724	7.823968	3.290724	2.703596	4.312661	3.155938	3.263919	3.020544	4.312661	3.541284	4.312661	2.970641
3.849093	4.034003	3.963237	3.382451	4.494262	7.578501	3.345604	2.718296	5.262342	3.108173	3.824925	2.86132	5.262342	4.101793	5.262342	4.478612

FP1	FP2	FC5	FC1	FC2	FC6	T7	C3	C4	T8	P3	P1	P2	P4	O1	O2	F
6.77244	5.198177	4.901128	4.691739	3.977873	5.537747	4.53735	5.241636	5.558295	3.411158	5.870315	3.861751	3.771363	3.546391	5.558295	3.668788	
7.082284	5.16186	5.205785	4.620282	4.186856	5.38914	4.698526	5.872968	6.194132	3.618645	6.201486	3.677695	3.953795	3.772896	6.194132	3.460981	
7.846531	5.651237	5.205254	4.44547	4.842453	5.958351	5.078725	5.974658	6.082182	3.661667	6.431812	3.666608	4.085225	4.474376	6.082182	3.443534	
6.47202	6.82478	7.256978	5.654207	6.619516	6.472074	6.221195	5.943083	8.944963	5.350794	5.952922	5.34242	8.944963	5.814876	8.944963	5.47963	
6.415529	6.598373	7.193852	5.494452	6.243077	6.367418	6.00788	5.895242	6.928104	5.189083	5.980619	5.053951	6.928104	5.389934	6.928104	5.044795	
7.176819	7.513876	8.028725	6.075178	6.943316	6.968462	6.897615	6.502179	7.623636	5.592375	6.197367	5.513002	7.623636	5.841639	7.623636	5.447113	
6.891127	6.965287	6.463721	5.874987	5.497982	8.699778	5.226609	4.819089	5.189667	4.783578	5.748308	3.928701	4.280322	5.331753	5.189667	4.54023	
7.543481	7.68683	7.106455	6.532675	6.074718	9.337692	5.968661	4.84284	5.715721	4.652007	5.985379	4.046936	4.461286	5.623443	5.715721	4.55438	
7.741361	7.934069	7.320306	7.042622	6.465878	9.576543	6.585155	5.05692	7.745588	4.922128	6.061691	4.060369	4.856596	5.462688	7.745588	4.609831	
3.984302	4.247251	4.465995	3.45266	4.45751	7.680142	4.323021	4.448719	7.65169	5.367821	5.388211	3.619986	7.65169	3.380097	7.65169	3.507609	
4.013489	4.817201	4.737701	3.551144	4.467525	7.314516	4.067183	4.926223	7.329736	5.120755	4.730603	3.250648	7.329736	3.363684	7.329736	3.255276	
4.487248	5.570072	5.580724	4.076687	5.380849	7.990603	4.522082	4.972605	11.03804	5.394417	5.080025	3.597051	11.03804	3.89956	11.03804	3.786856	
4.226557	5.436792	5.442315	3.867496	4.842572	7.865587	4.252086	5.040052	8.370697	5.196417	4.930152	3.370377	8.370697	3.632138	8.370697	3.280923	
4.459453	5.350736	5.606087	4.058463	4.95777	8.864829	4.507689	5.507809	7.575512	5.185729	5.235211	3.788376	7.575512	3.959192	7.575512	3.76592	
4.192739	4.69348	5.030073	3.716358	4.350074	8.686485	4.031801	4.991267	5.211452	5.335082	5.114458	3.545551	5.211452	3.565457	5.211452	3.854409	
4.206871	4.644534	4.72115	3.637935	4.235352	7.507535	4.046227	5.0015	5.381281	5.268159	5.040714	3.480031	5.381281	3.531979	5.381281	3.744028	
4.676743	5.457839	5.804705	4.171282	4.894103	8.071625	4.727719	5.388953	6.090154	5.659593	5.534127	4.181417	6.090154	4.281576	6.090154	4.344868	
4.368137	5.164673	5.427196	3.968776	4.533729	8.001784	4.43816	5.04078	5.096272	5.497784	5.40336	3.656958	5.096272	3.792962	5.096272	4.116165	
4.677539	5.528137	5.666381	4.232275	4.684264	7.898034	4.746533	5.195644	5.674811	5.420302	5.724668	3.7714	5.674811	4.285479	5.674811	5.607517	

FP1	FP2	FC5	FC1	FC2	FC6	T7	C3	C4	T8	P3	P1	P2	P4	O1	O2
6.531483	3.957054	4.076671	4.842105	3.204339	5.488153	4.043047	3.075875	6.537337	3.447053	3.647802	4.816099	3.801355	3.759134	6.537337	4.737293
6.881538	4.095054	4.559496	4.649786	3.143134	5.58129	4.196675	3.268251	7.567633	3.429854	4.139284	4.838654	3.919825	3.823798	7.567633	4.607375
7.472483	4.334982	4.405304	4.389576	3.985874	6.077879	4.398393	3.31026	6.813953	3.372591	4.19678	4.861538	4.006408	4.247472	6.813953	4.592902
5.884444	6.71432	7.031243	5.426642	6.173158	6.299709	5.916608	5.273323	11.05124	5.828723	5.141015	6.337225	11.05124	6.200654	11.05124	6.459594
5.877082	6.140421	6.881793	5.090963	5.939139	6.091775	5.594059	5.020607	8.665809	5.42692	4.932993	5.400824	8.665809	5.273573	8.665809	5.193816
6.284196	6.729262	7.252418	5.355933	6.260545	6.368819	5.935485	5.27699	8.836804	5.598242	5.159185	5.56176	8.836804	5.561939	8.836804	5.378587
6.566726	6.743331	5.870925	5.591222	5.048066	8.146727	5.195173	3.163903	6.495644	3.738302	3.482646	3.628969	3.636902	4.50747	6.495644	4.949944
7.23041	7.366809	6.663495	6.116447	5.631885	9.17934	5.260398	3.446468	7.356065	3.796935	3.67737	3.801004	4.025193	4.72109	7.356065	4.774025
7.302281	7.532592	6.911508	6.707421	6.019886	9.451541	7.144945	3.95969	9.552362	4.71743	4.792029	4.476579	4.199362	4.746415	9.552362	5.550511
5.162264	4.036778	4.60807	3.341439	6.799533	8.928385	5.939225	3.231805	8.192038	4.545987	5.273613	4.562742	8.192038	4.299867	8.192038	4.545553
3.637019	3.847831	4.047108	3.433016	6.908467	8.041393	4.640719	2.297767	7.133699	3.832447	3.968762	4.059301	7.133699	4.168942	7.133699	3.908118
4.510632	4.694629	4.786667	4.364439	7.705797	9.104514	4.452858	2.517291	14.40643	3.651383	4.364769	4.747887	14.40643	5.306064	14.40643	4.789646
3.915209	4.429443	4.294893	3.729238	6.647416	8.686467	3.886185	2.559927	9.905467	3.570784	4.538708	4.830893	9.905467	4.572156	9.905467	4.190758
4.800342	4.587431	4.575795	4.141707	7.335087	10.08916	4.323691	4.084877	8.804868	3.771219	5.198228	4.969834	8.804868	5.575339	8.804868	4.348385
5.339903	3.761178	4.017981	4.050528	6.052285	10.61169	3.87927	2.73931	6.348183	3.553964	4.018895	4.61624	6.348183	4.839961	6.348183	4.360432
5.56816	3.549752	3.616834	3.639038	6.151867	8.393258	4.280633	2.80858	6.536507	3.859993	3.964968	4.647556	6.536507	5.007983	6.536507	4.402841
5.601605	4.195498	4.662222	4.327611	6.663469	8.739022	4.673006	3.17229	6.841502	4.262477	4.604527	5.483565	6.841502	5.924957	6.841502	4.958473
5.136384	3.724413	4.118995	4.370875	5.972704	9.042295	4.212932	2.869857	5.392842	3.784165	4.085771	4.657221	5.392842	5.320544	5.392842	4.500259
5.276362	3.872277	4.024153	4.802876	6.351572	8.66576	4.390532	2.807142	6.70885	3.683999	5.090007	4.931823	6.70885	5.814284	6.70885	6.028267

FP1	FP2	FC5	FC1	FC2	FC6	T7	C3	C4	T8	P3	P1	P2	P4	O1	O2
0.96669	0.7659	0.83776	1.035386	0.813614	0.995826	0.892886	0.58493	1.177242	1.012239	0.621445	1.274132	1.015991	1.060492	1.177242	1.314284
0.973911	0.796768	0.877983	1.008413	0.753517	1.028833	0.894698	0.563353	1.231897	0.949616	0.669394	1.320032	0.997326	1.017046	1.231897	1.33764
0.953583	0.770277	0.849532	0.993162	0.824425	1.024239	0.867324	0.554501	1.107129	0.925519	0.653747	1.33525	0.983425	0.957461	1.107129	1.350166
0.90927	0.985477	0.965362	0.947796	0.933912	0.968471	0.945269	0.874882	1.252829	1.078418	0.849987	1.176147	1.252829	1.085245	1.252829	1.163957
0.916988	0.931635	0.957219	0.929228	0.951951	0.959172	0.932489	0.852293	1.263114	1.043913	0.826315	1.071029	1.263114	0.979183	1.263114	1.029179
0.941964	0.988641	0.957265	0.932445	0.957393	0.960174	0.919022	0.857919	1.185277	1.032447	0.873436	1.043904	1.185277	1.009952	1.185277	1.028041
0.954154	0.969092	0.907858	0.954272	0.919826	0.937054	0.991187	0.658299	1.24708	0.782376	0.606022	0.930339	0.853241	0.845649	1.24708	1.102082
0.959729	0.958669	0.938906	0.938102	0.927483	0.983293	0.883628	0.71297	1.289255	0.818009	0.615815	0.957897	0.908826	0.841991	1.289255	1.05578
0.944065	0.9499	0.944644	0.953014	0.930028	0.987807	1.09023	0.783428	1.243261	0.955453	0.78805	1.107007	0.870734	0.86933	1.243261	1.204223
1.277884	0.947007	1.033659	0.966284	1.522305	1.160389	1.369692	0.721198	1.068046	0.84421	0.982687	1.276921	1.068046	1.273094	1.068046	1.289205
0.911113	0.799701	0.860836	0.967132	1.538651	1.098436	1.145331	0.466505	0.969044	0.745841	0.837659	1.259608	0.969044	1.242244	0.969044	1.205688
0.982571	0.826999	0.84693	1.051622	1.45769	1.116876	0.972072	0.506967	1.38183	0.676923	0.846773	1.324469	1.38183	1.36002	1.38183	1.270265
0.926355	0.814998	0.788729	0.967279	1.384244	1.107541	0.913743	0.509195	1.19179	0.687107	0.91486	1.430755	1.19179	1.261174	1.19179	1.264936
1.066491	0.862387	0.822203	1.026	1.517142	1.15936	0.960354	0.703245	1.178639	0.731681	0.984137	1.328363	1.178639	1.429332	1.178639	1.163974
1.283521	0.807356	0.801624	1.090329	1.39628	1.217408	0.968749	0.548514	1.208654	0.669806	0.783032	1.324861	1.208654	1.369092	1.208654	1.145043
1.315808	0.772636	0.777247	1.011675	1.455655	1.116976	1.068231	0.563786	1.210353	0.731785	0.785562	1.338233	1.210353	1.417252	1.210353	1.176523
1.192653	0.767093	0.807164	1.040577	1.36575	1.077824	0.989874	0.588167	1.111483	0.752429	0.827498	1.326971	1.111483	1.387186	1.111483	1.148365
1.176169	0.724348	0.762299	1.09919	1.31119	1.119944	0.948232	0.569654	1.052041	0.687065	0.755655	1.286485	1.052041	1.407591	1.052041	1.10038
1.139575	0.708803	0.712171	1.132073	1.366158	1.093469	0.923577	0.541048	1.2009	0.680868	0.88499	1.349055	1.2009	1.375556	1.2009	1.100236

Alpha FPR	Alpha FCR	Alpha FR-F	Beta diff fr	Theta/beta	GSR	Pulse	Response
-1.64005	-0.4343	-2.07435	-3.01624	-0.8457	3.256904	88.98817	0
-2.15666	-0.09861	-2.25527	-2.99218	-0.84193	3.134836	86.38806	0
-2.27822	0.962743	-1.31547	-1.278	-0.8021	3.020261	77.92497	1
0.558795	-0.26314	0.295651	2.186332	-0.53318	3.608751	74.92	0
0.554688	0.704018	1.258706	1.478829	-0.57829	3.517526	81.44245	1
0.519283	-0.35505	0.164234	3.695616	-0.51484	3.709054	78.2187	1
0.175228	1.739777	1.915005	3.433009	-0.59507	3.772949	80.88915	1
-0.11297	2.140247	2.027277	5.369201	-0.5516	3.714419	79.78389	1
0.161926	2.05699	2.218916	6.246451	-0.55331	4.033913	74.78375	1
-0.25628	6.136823	5.880544	-0.39483	0.835084	3.477013	70.50133	0
0.299807	5.392069	5.691875	-0.18049	0.64217	3.575798	66.90137	0
0.99394	5.643066	6.637006	2.161795	0.606947	3.697624	67.43743	0
0.971279	5.300197	6.271475	1.365223	0.512164	3.67745	71.61225	0
0.843968	5.738068	6.582036	2.393607	0.744271	3.880917	74.67287	0
-0.2025	6.594008	6.391507	0.95031	0.626426	3.832999	74.05679	0
-0.55022	5.113816	4.563598	-0.39231	0.550649	3.698904	71.66621	0
0.032536	4.929358	4.961894	1.643838	0.416049	4.077651	93.7853	0
0.027387	4.776845	4.804232	0.920457	0.360863	4.258503	81.67065	1
0.18491	4.727074	4.911984	1.330707	0.373811	4.867335	79.16364	1

The findings of our research show that 3 out of 4 subjects succeeded to obtain a low fear level using our system. Also, we highlight the limits of our approach: the instability of the sensor devices, a lot of noises and errors were introduced in data, which we had to clean. Also, we must mention the discomfort of the EEG device, which led to its abandonment.

We extended our work presented in [9] by defining a ML-based decision support and increasing the number of participants to 8 in [10] (Figure 3.9).

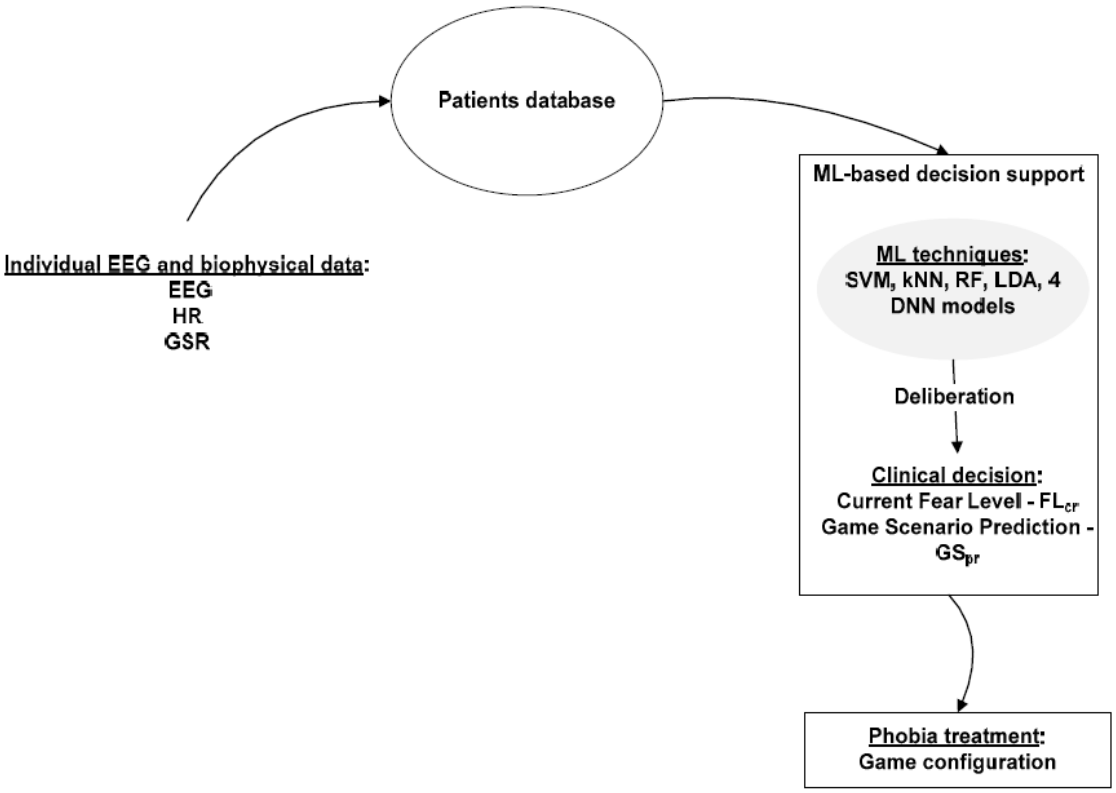

Figure 3.9. ML-based decision support for VRET [10]

In [10], we added kNN, SVM, RF, LDA classifiers for both level of fear and the game level that should be played next and two modalities to calculate the accuracies of the classifiers: an user-dependent and an user-independent. In the user independent modality each classifier was trained using the data of the other subjects and we applied the trained model on the data of the tested user to calculate cross-validation and test accuracies. In the user-dependent modality, for each subject, each classifier has been trained, cross-validated and tested on his/her own data.

We obtained the highest cross-validation and test accuracies for fear and game level classifiers presented in Table 3.29.

Table 3.29 The highest accuracy for fear and game level prediction [10]

Method	Fear level prediction				
	2-choice scale		4-choice scale		11-choice scale
	Cross-validation	Test	Cross-validation	Test	Cross-validation
User-indepedent	kNN 99.5% RF 99.25%	DNN_Model_4 79.12%	kNN 99% RF 99%	kNN 52.75%	kNN 98.25% RF 99%
User-dependent	kNN 99.5% RF 99.75%	SVM 89.5%	kNN 99% RF 99.25%	SVM 42.5%	kNN 98.25% RF 99%
	Game level prediction				
	2-choice scale		4-choice scale		11-choice scale
	Cross-validation	Test	Cross-validation	Test	Cross-validation
User-indepedent	RF 99.75%	-	RF 100%	-	RF 100%
User-dependent	RF 99.75%	-	RF 99.75%	-	RF 100%

With RF algorithm, we calculated the feature importance for each feature on the outputs of the classifiers in both methods, user-independent and user dependent. We found that GSR, HR and the beta waves had the most significant influence in fear level prediction as we noted in state-of-the-art literature [78], [79]. The most relevant features for fear predictions are shown in Tables 3.30 – 3.31.

Table 3.30. Feature importance for fear prediction, user - independent modality [10]

2-choice scale		4-choice scale		11-choice scale	
Feature	Feature importance	Feature	Feature importance	Feature	Feature importance
GSR	0.41	GSR	0.45	GSR	0.49
HR	0.28	HR	0.28	HR	0.24
B_C3	0.15	B_FC6	0.15	B_FC6	0.14

B_P3	0.13	B_C3	0.13	B_FC5	0.12
B_FC2	0.13	B_FC2	0.12	B_C3	0.12
B_FC6	0.13	B_FP1	0.12	B_FC2	0.12
B_FP2	0.12	B_P3	0.12	B_P3	0.11
A_FC6	0.12	T_FC6	0.12	T_FC6	0.11
B_C4	0.10	B_O1	0.11	B_FP1	0.10
B_FC5	0.10	B_FC5	0.11	A_FC6	0.10
B_FP1	0.09	B_T8	0.09	B_T8	0.10
T_FC6	0.08	B_P2	0.09	B_O1	0.08
A_FP1	0.08	B_FC1	0.08	A_FP1	0.08
A_FP2	0.08	A_FP1	0.08	B_P2	0.08
B_T8	0.08	A_FC6	0.08	T_FP1	0.08

Table 3.31. Feature importance for fear prediction, user - dependent modality [10]

2-choice scale		4-choice scale		11-choice scale	
Feature	Feature importance	Feature	Feature importance	Feature	Feature importance
GSR	0.40	GSR	0.46	GSR	0.48
HR	0.25	HR	0.32	HR	0.27
B_FC2	0.22	B_FC6	0.17	B_FP1	0.14
B_C4	0.15	B_FC2	0.16	A_FC6	0.14
B_FC6	0.14	B_P2	0.12	B_FC2	0.14
A_FP1	0.14	B_FP1	0.12	B_FC6	0.13
B_P2	0.13	T_FC6	0.11	T_FC6	0.12
A_FC6	0.12	B_O1	0.10	B_O1	0.12
B_FP1	0.10	A_FC6	0.10	A_FP1	0.11
B_O2	0.10	A_FP1	0.10	B_FC5	0.11
T_P2	0.08	B_P3	0.10	B_P2	0.10
T_FC6	0.08	B_C4	0.09	B_P3	0.10
B_O1	0.08	B_FC5	0.09	B_C3	0.09
B_C3	0.08	A_P2	0.08	B_T8	0.09
B_P3	0.08	B_C3	0.08	B_C4	0.08

The human-centered approach for VRET systems

In [6], we stressed that the usage of VR-based applications with AI-based algorithm embedded for phobia treatment requires a human-centered approach for machine learning models, ensuring patient safety was the tenet of developing such applications. The proposed human-centered approach for VRET systems development consisted of 4 stages supervised by the therapist:

- Stage I Gameplay and biophysical data recording supervised by the therapist.
- Stage II Biophysical data processing supervised by the therapist.
- Stage III ML-based fear level classification supervised by the therapist.
- Stage IV ML-based VR scenario selection supervised by the therapist.

For machine learning we adopted the Interactive Machine Learning (iML) presented in [127]. The opaque-box view for ML models is replaced by a glass-box, in which the human experts seen as agents are involved in all steps of their development: data, training, testing, deployment. iML can interact with agents and optimize their learning behaviors.

We defined in [7] a methodology inspired by the Human-Centered Distributed Information Design (HCDID) model ([73]) and from the four-layer model of human Needs and Aspirations for application in a Design and Innovation process (NADI) model ([74]) for designing and developing a VRET system (Figure 3.10).

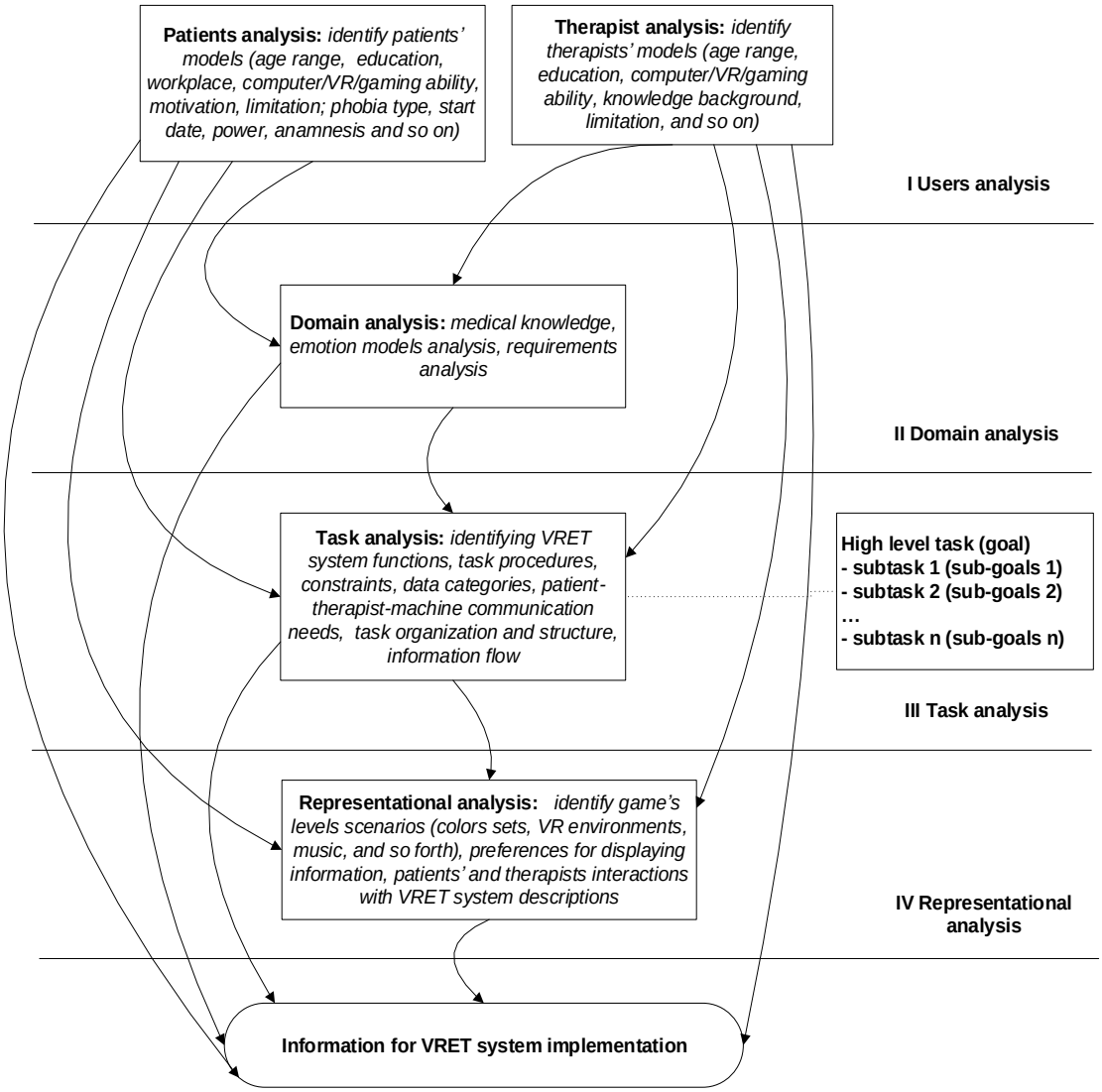

Figure 3.10. Layers-based analysis for designing a human-centered VRET systems [7]

Our methodology is comprised of 4 levels:

- Level 1 – deals with users' analysis, including patients and therapists. At this stage there were identified and recorded features of users' patterns.
- Level 2 – is concerned with the system analysis requirements, emotion models and medical knowledge.
- Level 3 – performs task analysis. Each task had a hierarchical structure according to the goals and sub-goals of the system.
- Level 4 - deals with representational analysis (for example, the patients' and therapists' preferences for colours or sounds, for a certain game, for urban or natural landscapes, for certain technologies and so on).

This methodology was used in development of the second game for phobia treatment [7], [8].

A holonic-based architecture for VRET systems

We proposed in [11] a holonic architecture for VRET systems. The term holon coined by Arthur Koestler in 1967 to distinguish the relation between wholes and parts refers to a component which can act intelligently [80]. Later, the concept was used in manufacturing systems and the holon is seen as a block with autonomy, cooperation, self-organization and re-configurable having two parts: an information processing part and a physical part [81]. A hierarchy of holons is called holarchy and a holon itself can be a holarchy. Thus, holons can have a recursive structure. Our approach was inspired by previous research of the author of this thesis presented in [82].

Considering the design a VRET system capable of treating multiple phobias, we defined an architecture based on the holon, called PhoVRET (Figure 3.11.).

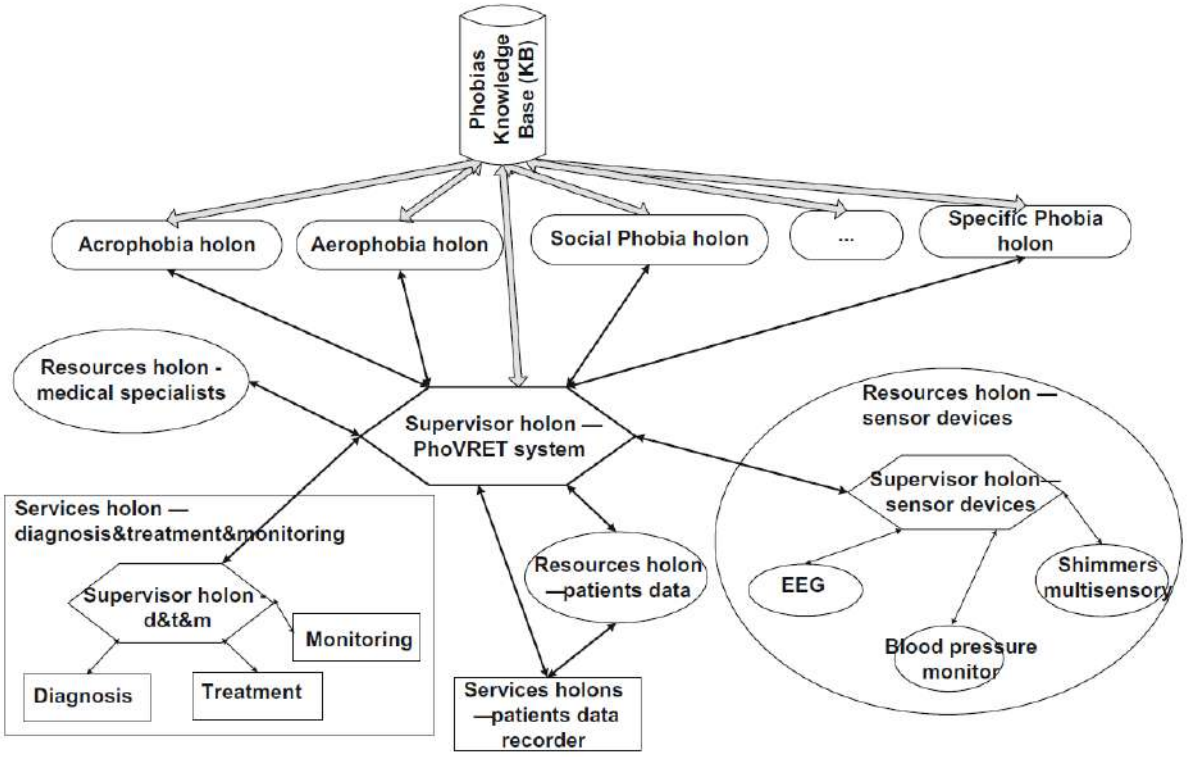

Figure 3.11. Holonic architecture for VRET systems [11]

Four types of holons are defined: supervisor holons (hexagon shapes), services holons (rectangle shapes), resources holons (ellipse shapes), specific phobia holons (rounded rectangle shapes). The architecture consists of multiple holarchies coordinated by the supervisor holons. The main supervisor holon coordinates all PhoVRET holons through the supervisors of the holarchies. A service type holon provides services, e.g. the Diagnosis & Treatment & Monitoring service holon deals with phobias diagnosis, treatment and

monitoring. The specific phobias holons Acrophobia Holon, Aerophobia Holon, Social Phobia Holon, etc. guide to the treatment of specific phobias. The primary resources of the system (patients' data, sensor devices, physicians, etc.) are managed by the resources holons.

All holons are sociable both with themselves and with humans. The cooperation between holons and holons and humans is achieved using the cooperation domain-based approach. A cooperation domain is assigned to a group of holons.

As in [83], we considered the structure of a holon as being made up of the information processing part (containing three modules - interholon interface, human interface, and the holon's kernel in charge of decision making) and, where needed, the physical processing part (hardware and controller of the hardware).

For acrophobia, the prototype of the VRET system is shown in Figure 3.12.

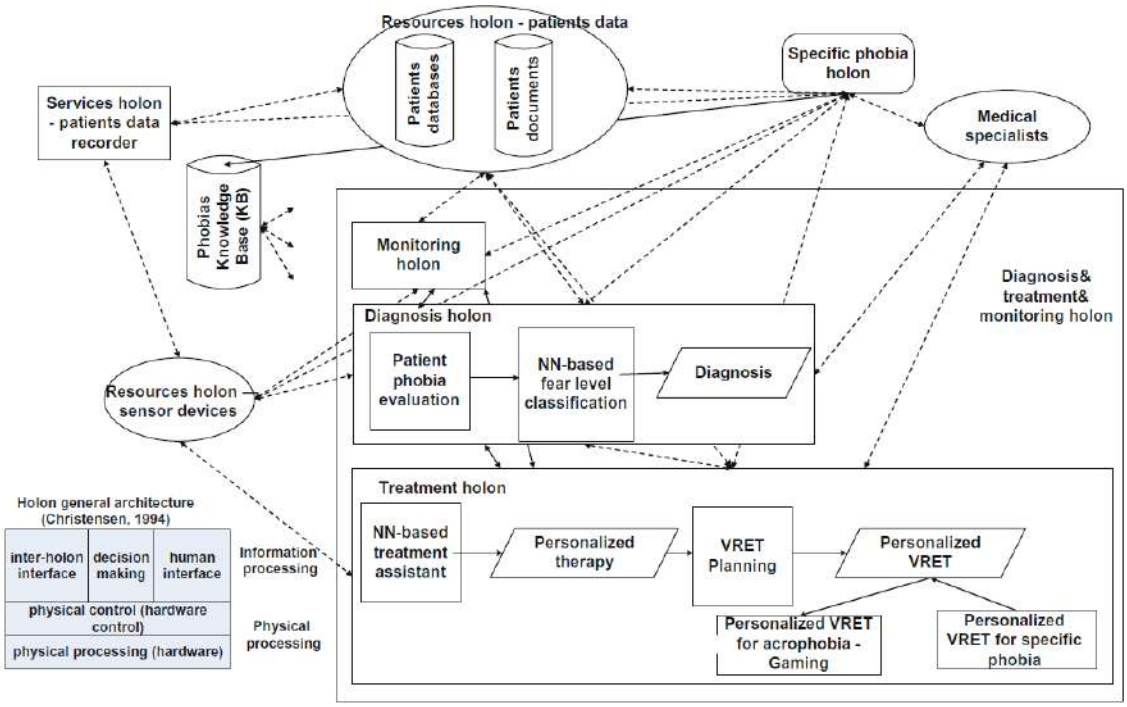

Figure 3.12. A prototype of the VRET system based on holons [11]

A holon-based architecture is difficult to be implemented, but at the same time a system with this type of architecture can be very easily maintained. Our implementations have shown that the holon-based architecture is feasible for a VRET system.

Challenges for integrating AER in VRET systems

Two main challenges for integrating AER in VRET systems were identified: one related to data acquisition and one to resource consumption. The data acquisition-related concerns

refer to the artefacts identifying (if they existed) and removing them from the data. We covered the subject in [8]. We focused on the ways in which biosignals should be measured in the VR environments to obtain valid data [8].

We defined "artefacts as any misleading or confusing alterations in physiological data that appear as a result of external action such as head, hand or body movements, being unrelated to the emotional effects that specific stimuli or the object under observation exert upon the user" [8]. To recognize the patterns of artefacts when acquiring signals during a VR game playing, we proposed a method consisting of "aligning the data segments corresponding to the moments when the users performed head/hand/body movements with the artefact signals recorded during a reference procedure" [8].

To validate our methods, a new VR game for acrophobia therapy was developed. 5 participants played the game several times and we captured their physiological signals (GSR, HR and RR). The game's scenes consisted of a mountain landscape with peaks, valleys, forests and a lake. The players took a ride using a cable car from the lowest height point located in a valley to the highest peak of the mountain. The scene was rendered via the HTC Vive Head Mounted Display and the interaction was ensured by controllers. Heart Rate and Galvanic Skin Response were measured by the Shimmer3 GSR+ Unit (https://www.shimmersensing.com/product/shimmer3-gsr-unit/) and the Respiration Rate was computed according to the distance between two HTC Vive trackers.

The method to artefacts' recognizing consists in three steps:

- Step I Reference artefact measurement.
- Step II Artefact measurement during gameplay.
- Step III Artefact matching evaluation.

In Step I - a set of measurements were done, both baseline and reference artefacts for each user. The users had to perform several movements (of head, hand, and body) to be recorded artefacts patterns. In our experimental study, we recorded and saved GSR, HR and RR to be mapped onto the same type of data recorded during the gameplay sessions.

In Step II, during the gameplay sessions - the users performed artefact-like tasks (movements) at specified points. The biophysical data and the timestamps when the movements have been effectuated are recorded. So, we assure the alignment of the data to perform a point-to-point comparison.

In Step III - we made an alignment and a correspondence between the biophysical data recorded as reference artefact measurements at Step I with the biophysical data recorded during the gameplay sessions at Step II. We verified the match between reference artefacts

segments and data recorded in the game segments by computing the bias or Mean Absolute Error (MAE) or the Mean Absolute Percentage Error (MAPE) for GSR and HR.

"A segment of data is an artefact if there is a difference between adjacent data segments recorded before and after few seconds from the moment marked with artefact-like task and the reference artefact" [8].

The screenshots with the code for computing the values for MAE, MAPE, MAPE previous and after the moment marked with artefact-like task are provided below.

- Code for computing Mean Absolute Error (MAE) for GSR and HR

```
for (int k1 = posStart, k2 = 0; k1 < posStop; k1++, k2++)
{
error = Math.Abs(listGSRGame[k1] - listGSRArtefact[k2]);
sumErrorsGSR = sumErrorsGSR + error;
error = Math.Abs(listHRGame[k1] - listHRArtefact[k2]);
sumErrorsHR = sumErrorsHR + error;
}
errorGSR = sumErrorsGSR / listGSRArtefact.Count;
errorHR = sumErrorsHR / listHRArtefact.Count;
```

- Code for computing Mean Absolute Percentage Error (MAPE) for GSR and HR

```
for (int k1 = posStart, k2 = 0; k1 < posStop; k1++, k2++)
{
error = Math.Abs(listGSRGame[k1] - listGSRArtefact[k2])/Math.Abs(listGSRGame[k1]);
sumErrorsGSR = sumErrorsGSR + error;
error = Math.Abs(listHRGame[k1] - listHRArtefact[k2])/Math.Abs(listHRGame[k1]);
sumErrorsHR = sumErrorsHR + error;
}
errorGSR = sumErrorsGSR / listGSRArtefact.Count*100;
errorHR = sumErrorsHR / listHRArtefact.Count*100;
```

- Code for computing Mean Absolute Percentage Error before few seconds from the moment marked with artefact-like task for GSR and HR

```
for (int k1 = posStart - step* listGSRArtefact.Count, k2 = 0; k1 < posStop - step*
listGSRArtefact.Count; k1++, k2++)
```

```
{
error = Math.Abs(listGSRGame[k1] - listGSRArtefact[k2]) /
Math.Abs(listGSRGame[k1]);
sumErrorsGSR = sumErrorsGSR + error;
error = Math.Abs(listHRGame[k1] - listHRArtefact[k2]) / Math.Abs(listHRGame[k1]);
sumErrorsHR = sumErrorsHR + error;
}
errorGSR = sumErrorsGSR / listGSRArtefact.Count * 100;
errorHR = sumErrorsHR / listHRArtefact.Count * 100;
```

- Code for computing Mean Absolute Percentage Error after few seconds from the moment marked with artefact-like task for GSR and HR

```
for (int k1 = posStart + step* listGSRArtefact.Count, k2 = 0; k1 < posStop + step*
listGSRArtefact.Count; k1++, k2++)
{
error = Math.Abs(listGSRGame[k1] - listGSRArtefact[k2]) /
Math.Abs(listGSRGame[k1]);
sumErrorsGSR = sumErrorsGSR + error;
error = Math.Abs(listHRGame[k1] - listHRArtefact[k2]) / Math.Abs(listHRGame[k1]);
sumErrorsHR = sumErrorsHR + error;
}
errorGSR = sumErrorsGSR / listGSRArtefact.Count * 100;
errorHR = sumErrorsHR / listHRArtefact.Count * 100;
```

In our experiments, we tested the following artefacts: deep breath, head movement to the left, head movement to the right, head movement up, head movement down, click with the right hand on the HTC Vive controller, right hand raise. 5 participants aged 24-50 were volunteers in our experiment. The experiment was performed in a controlled environment (quiet room with temperature 25 degree Celsius).

The experiment consisted in 2 phases:

- Phase 1 Reference artefact measurements were performed for each user. So, we have obtained patterns for artefacts.
- Phase 2 The participants played the VR-based game, having 10 stops during the ride. At each stop, the players performed in a time of 5 seconds the tasks associated with

artefacts: a deep breathing, a head movement to the left/right/up/down, a click on the controller and a raising of the hand. The HR, GSR, RR and timestamps were recorded.

The software used in this experiment was developed in-house. Extracts from the results obtained for a user are shown in Table 3.28.

Table 3.28. Example of results for MAE, MAPE, MAPE before and after

MAE
1 Respiration 1 GSR=0.315750915765579 HR=3.91812865497076 posStart = 1778 posStop = 2120
1 Respiration 2 1.07841002314999 3.810888252149 posStart = 1778 posStop = 2127
1 LeftHeadMovement 1 0.236734693888444 3.55393586005831 posStart = 2293 posStop = 2636
1 LeftHeadMovement 2 0.866596756225551 4.50340136054422 posStart = 2293 posStop = 2587
1 RightHeadMovement 1 0.187866634577297 2.54014598540146 posStart = 2805 posStop = 3079
1 RightHeadMovement 2 0.71969172518437 1.29357798165138 posStart = 2805 posStop = 3132
1 DownHeadMovement 1 0.238273280509989 3.8207171314741 posStart = 3314 posStop = 3565
1 DownHeadMovement 2 0.472816072808899 4.50757575757576 posStart = 3314 posStop = 3578
1 UpHeadMovement 1 0.304287996867305 2.12166172106825 posStart = 3829 posStop = 4166
1 UpHeadMovement 2 0.679286486199318 2.3448275862069 posStart = 3829 posStop = 4090
1 RightHandClick 1 0.22237331013872 8.56115107913669 posStart = 4344 posStop = 4483
1 RightHandClick 2 2.27399071601873 5.63636363636364 posStart = 4344 posStop = 4520
1 UpRightHand 1 0.241949504334343 7.46917808219178 posStart = 4852 posStop = 5144
1 UpRightHand 2 2.1237696808027 8.13913043478261 posStart = 4852 posStop =

5082

MAPE
1 Respiration 1 8.91182782920685 4.82160645271041 posStart = 1778 posStop = 2120
1 Respiration 2 30.6664729829737 4.79313019051415 posStart = 1778 posStop = 2127
1 LeftHeadMovement 1 7.02005867557086 4.45994242689304 posStart = 2293 posStop = 2636
1 LeftHeadMovement 2 25.6445219513772 5.64795207075351 posStart = 2293 posStop = 2587
1 RightHeadMovement 1 5.78805305159316 3.30972897322338 posStart = 2805 posStop = 3079
1 RightHeadMovement 2 22.2135036003045 1.66060493920209 posStart = 2805 posStop = 3132
1 DownHeadMovement 1 7.33918076728729 4.81538922425609 posStart = 3314 posStop = 3565
1 DownHeadMovement 2 14.5759266790874 5.56554741851668 posStart = 3314 posStop = 3578
1 UpHeadMovement 1 11.8817213580432 2.66807822594173 posStart = 3829 posStop = 4166
1 UpHeadMovement 2 24.4240284055912 2.97976225562432 posStart = 3829 posStop = 4090
1 RightHandClick 1 7.24958477911861 10.1918465227818 posStart = 4344 posStop = 4483
1 RightHandClick 2 73.8193329631399 6.71396504729837 posStart = 4344 posStop = 4520
1 UpRightHand 1 6.54895622988269 8.65404429170641 posStart = 4852 posStop = 5144
1 UpRightHand 2 57.6426117585725 9.33593854942368 posStart = 4852 posStop = 5082

MAPE after_1
1 Respiration 1 6.62114617405325 5.57270437188454
1 Respiration 2 33.9910238379181 4.48343212908636
1 LeftHeadMovement 1 4.34301564439196 6.57202343608583
1 LeftHeadMovement 2 28.960124872332 5.3377239848759
1 RightHeadMovement 1 5.86482790544933 5.93733189575623
1 RightHeadMovement 2 21.4714367648972 4.56684171974895
1 DownHeadMovement 1 4.91701607236821 5.50755725320623
1 DownHeadMovement 2 17.6598352591281 2.26356739514634
1 UpHeadMovement 1 10.0586631519476 6.53172248127739
1 UpHeadMovement 2 43.4940781587203 10.9377850757161
1 RightHandClick 1 2.77665268236384 7.65451131158565
1 RightHandClick 2 62.1260873948667 2.94417451590447
1 UpRightHand 1 8.73441203898137 9.12482195689828
1 UpRightHand 2 59.5145318376213 7.72043600048717

MAPE after_2
1 Respiration 1 3.02693682922576 4.2284433792069
1 Respiration 2 39.156534500154 3.20484079592382
1 LeftHeadMovement 1 3.22460166029285 9.05281770358699
1 LeftHeadMovement 2 31.505918665577 6.48315718791185
1 RightHeadMovement 1 5.29445349095248 5.06748305502909
1 RightHeadMovement 2 24.3014997836889 1.69387752680847
1 DownHeadMovement 1 6.44350849667536 4.29888484868564
1 DownHeadMovement 2 31.4009102314032 4.23401973401972
1 UpHeadMovement 1 11.7462305833746 2.76849556044238
1 UpHeadMovement 2 14.2890083471291 9.70625798212002
1 RightHandClick 1 4.86672406581129 5.11852770299178
1 RightHandClick 2 57.6826862866468 3.0433042695701
1 UpRightHand 1 10.4058704364235 7.39910799289132
1 UpRightHand 2 62.9131170073817 11.8529009942252

MAPE after_3
1 Respiration 1 0.923223381461565 3.38584922770904
1 Respiration 2 42.5166160142029 4.82631885672077
1 LeftHeadMovement 1 2.88554199570595 5.70068883185696
1 LeftHeadMovement 2 30.051242794142 4.7168226159443
1 RightHeadMovement 1 2.79088327543697 3.06646869689513
1 RightHeadMovement 2 38.3347925223248 3.37814374511621
1 DownHeadMovement 1 22.181359810384 4.83523780336529
1 DownHeadMovement 2 45.832329195864 11.4628427128428
1 UpHeadMovement 1 20.5297644257908 7.58427332054121
1 UpHeadMovement 2 6.68587070470393 5.46698393183028
1 RightHandClick 1 5.50060450999563 5.2952861062681
1 RightHandClick 2 44.4045537184962 9.91225761069386
1 UpRightHand 1 11.5106458137365 4.39318952214388
1 UpRightHand 2 66.5994564166515 6.83779934582265

MAPE previous_1
1 Respiration 1 4.66462468740111 8.54844595498357
1 Respiration 2 49.5548274076336 2.49269737493256
1 LeftHeadMovement 1 11.265503439817 5.82260907462962
1 LeftHeadMovement 2 20.5984164428105 4.24185436682207
1 RightHeadMovement 1 7.53326260095634 5.74199052781762
1 RightHeadMovement 2 19.4375112243698 4.03181322993191
1 DownHeadMovement 1 7.20159543295731 3.88917348593632
1 DownHeadMovement 2 14.7175870002865 4.02653154984934
1 UpHeadMovement 1 7.91411547855986 3.63556575878925
1 UpHeadMovement 2 13.5704323219003 2.49014895112898
1 RightHandClick 1 26.1387100768134 10.1918465227818
1 RightHandClick 2 107.337892903865 6.79112554112554
1 UpRightHand 1 15.9739915606448 2.32913310842333
1 UpRightHand 2 73.5286139699928 2.57974307231325

MAPE previous_2
1 Respiration 1 4.8199161849846 4.58333333333331
1 Respiration 2 50.2093972580632 3.39054808610681
1 LeftHeadMovement 1 6.75054401640385 11.2429587154682
1 LeftHeadMovement 2 23.8473531534174 3.07339716708602
1 RightHeadMovement 1 9.99523772247836 9.03326552729484
1 RightHeadMovement 2 15.2877888402585 7.51444104655117
1 DownHeadMovement 1 7.36347545408528 5.82273168243211
1 DownHeadMovement 2 14.2883165471503 2.43324434233526
1 UpHeadMovement 1 10.3657004006746 3.45059310081639
1 UpHeadMovement 2 10.5961031480705 5.68640025036912
1 RightHandClick 1 36.7854200243231 9.84135768308428
1 RightHandClick 2 118.460643635443 6.75314269064268
1 UpRightHand 1 30.2568376101639 6.15018965475586
1 UpRightHand 2 83.2727966950111 6.03653860900237

MAPE previous_3
1 Respiration 1 3.34953034745701 5.49366375126784
1 Respiration 2 47.9689814738087 3.11568667930267
1 LeftHeadMovement 1 2.59207931920947 7.53990057561486
1 LeftHeadMovement 2 39.0829317297052 6.26739625163995
1 RightHeadMovement 1 13.4773764272826 8.51970340940592
1 RightHeadMovement 2 11.4416527673288 7.20612447226659
1 DownHeadMovement 1 8.91867750897962 2.64703745906516
1 DownHeadMovement 2 12.2771205272873 4.94689421108739
1 UpHeadMovement 1 9.69795769425905 3.00277686889779
1 UpHeadMovement 2 10.8642839896896 3.46580355539901
1 RightHandClick 1 17.8463669106151 3.26581811473897
1 RightHandClick 2 72.0627112669739 4.8423895923896
1 UpRightHand 1 55.0832777386557 5.98499673842139
1 UpRightHand 2 119.75622343193 7.62422360248449

We computed the average value for MAE, MAPE, MAPE after, MAPE before for GSR and HR and obtained the results from Table 3.29.-3.31.

Table 3.29. The average GSR MAPE [8]

Artefact type	GSR MAPE (%)				
	GSR -2 steps	GSR -1 step	GSR aligned time stamps	GSR +1 step	GSR +2 steps
RESPIRATION	28.26856386	28.13125142	28.3442362	29.36369501	29.65567029
LEFT HEAD	27.58241056	28.18084292	29.2564398	29.19821758	29.06203135
RIGHT HEAD	28.29350035	28.32949923	28.3200534	28.35180869	28.03708074
DOWN HEAD	27.81568491	27.77749042	27.62714529	27.33337573	27.5452365
UP HEAD	27.92170684	27.57325531	27.79352328	27.58577201	27.03417432
CLICK	31.81846513	31.28737745	30.76154053	30.69721127	30.79128381
UP HAND	33.74056947	33.83259569	32.68694971	32.58406055	32.81968315
AVERAGE FOR ALL ARTEFACTS	29.34870016	29.30175892	**29.25569832**	29.30202012	29.27788002

Table 3.30. The average HR MAPE [8]

Artefact type	HR MAPE (%)				
	HR -2 steps	HR -1 step	HR aligned timestamps	HR +1 step	HR +2 steps
RESPIRATION	7.169156332	7.162239718	7.209832312	6.605203634	8.723910664
LEFT HEAD	10.02081962	9.836236208	8.776661229	10.15688102	11.64451186
RIGHT HEAD	11.0489636	10.06476219	10.88219041	9.977128487	10.54933027
DOWN HEAD	11.04079454	9.192678831	8.362914572	8.955357568	9.487560067
UP HEAD	9.251944032	8.775189409	8.544770864	9.221602547	8.498722493
CLICK	8.299299978	8.376145477	8.285190912	7.644017162	8.463458326
UP HAND	8.618777591	7.483971872	9.087532661	8.317930388	7.314622975
AVERAGE FOR ALL ARTEFACTS	9.349965098	8.698746244	**8.735584708**	8.696874401	9.240302379

Table 3.31. MAE for aligned timestamps [8]

Artefact type	MAE for aligned timestamps	
	GSR (microSiemens)	HR (beats per minute)
RESPIRATION	0.51	5.33
LEFT HEAD	0.51	6.46
RIGHT HEAD	0.48	7.80
DOWN HEAD	0.46	6.09
UP HEAD	0.45	6.13
CLICK	0.55	6.20
UP HAND	0.60	6.64
AVERAGE FOR ALL ARTEFACTS	**0.5**	**6.37**

By analysing the data, we concluded that the bias is lower, but not significantly, on the aligned data segments than on the segments before and after the moments with artefacts. The deep breath artefact had a stronger influence than the other artefacts. Related to RR, the measured values were identical in both situations, the reference and gameplay session.

A specially designed protocol for the acquisition of biophysical signals in VR environments for phobia treatment was proposed in [12]. The protocol also aimed to improve signal processing to extract the most relevant features for anxiety estimation.

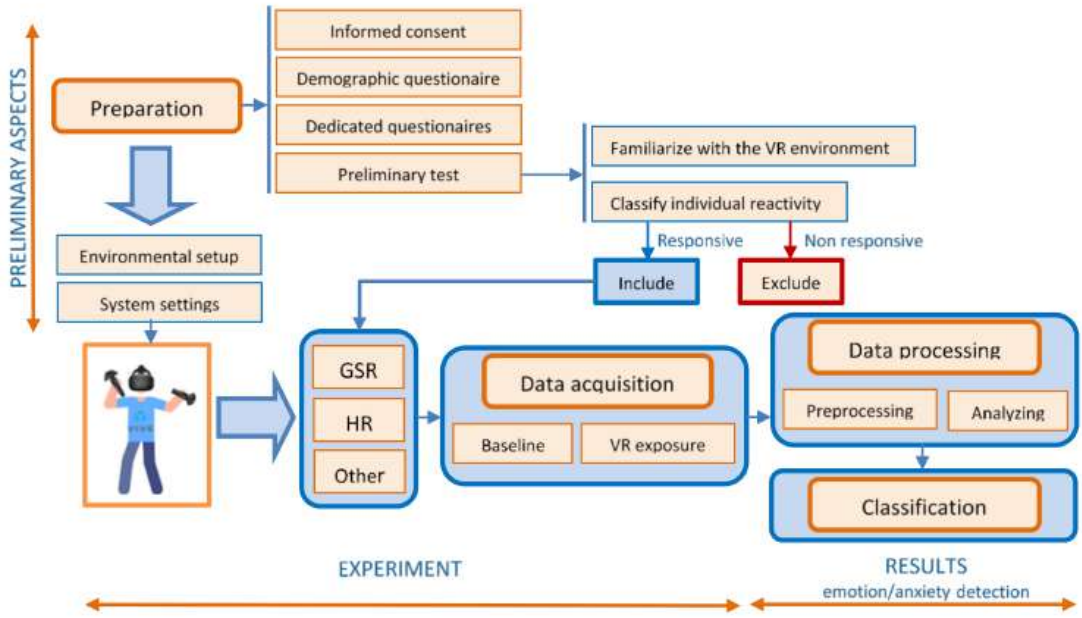

Figure 3.11. Protocol for biophysical signals acquisition in VR environments and emotion classification [12]

The stage of preparation involved:

- Obtaining the consent from the participants.
- Meeting the participants and performing a preliminary psychometric evaluation.
- Familiarizing the participants with the VR equipment.
- Assessing the individual reactivity and excluding the individuals without reaction.

The stage of data acquisition consisted of:

- Setting an appropriate environment (quiet, optimal ambient light, noise-free, constant temperature and humidity, the same period of day for conducting the experiment).
- Setting and calibration all devices.
- Recording data in baseline mode.

The stage of data processing was composed of:

- Data preprocessing with data filtering, downsampling, normalization and standardization.
- Data analyzing with data deconvolution, feature extraction.

The last stage of the emotion classification involved obtaining the intensity of the emotion using trained ML models (classical or convolutional neural networks).

The data acquisition protocol was validated in the laboratory by performing an experiment on 7 subjects. The subjects were exposed to various height levels (fear stimuli) in VR-situations. Our interest was in extracting features and to obtain the best combinations of them to be used in the emotion (anxiety) estimation function. We considered 32 features (11 types of features from EDA and 5 types from HR) and made all possible combinations from 1 feature to 7 features from the 32 features. We used a regression analysis to estimate the relationship between the intensity of the anxiety and features extracted from EDA/HRV signals.

We obtained a Sum Squared Error Estimation decreasing as follows: 3.827 (in the case of 1 feature), 3.059 (2 features), 2.663 (3 features), 2.041 (4 features), 1.656 (5 features), 1.286 (6 features), and 1.031 (7 features) [12].

Chapter 4. Building Creative Groups in Collaborative Learning Using Artificial Intelligence Techniques

This chapter is based on the articles published by the author of this habilitation thesis as co-author in [20], [21], [22], [23]. The period we refer to in this chapter is between 2013-2018, during which the author conducted research related to intelligent computational models for building creative teams. The research question originated from a real situation, encountered in the educational process, namely, how we build teams of CS students capable of providing creative software solutions to real-world problems.

Findings

- A Q-learning algorithm - based method to build in an optimal way the most creative learning groups.
- A multiagent system to build creative students' group.
- A Bayes classifiers-based model to build the most creative students' group.

The Importance of the topic

Collaboration is a skill often required for graduates in real-life scenarios and has been a permanent concern designing the learning process to acquire it. The development of creativity and critical thinking are also a requirement, and the learning process must be focused on these concepts as well. Collaborative learning (CL) is defined in [84] as "groups of learners working together to solve a problem, complete a task, or create a product". Computer supported collaborative learning (CSCL) emerged in the 1990s and refers to ways in which people can learn together with the help of computers [85]. It was a reaction to the singularity of the individual in front of the computer when taking online courses.

Highly debated in specialized literature, creativity is viewed by Sternberg et al. (2005) as "the ability to produce work that is novel (i.e., original, unexpected), high in quality, and appropriate (useful)" [86]. The investment theory of creativity offers an interesting perspective on an individual considered creative "are ones who are willing and able to metaphorically buy low and sell high in the realm of ideas" [87], [88].

Group creativity shows "the social nature of creative act" [89]. Group creativity is not just a simple reunion of the creativity of the individuals in the group but represents the creativity

resulting from the interaction of different individuals with varied backgrounds, cultures, abilities, and knowledge [20]. Creativity in the context of collaborative learning emphasizes learning because of interactions and collaborations between people with the goal of increasing creativity at both the individual and group levels.

The research question that guided our interests was: Can student group's creativity be enhanced by providing contextual instructional environments and organizing individuals into appropriate groups using artificial intelligence techniques?

To model groups' creativity, we investigated more references, of which we mention the most relevant ones. The Componential Model of Creativity proposed by Amabile for individual creativity and later extended to team creativity [90], [91], [92], [93], which includes both individual and non-individual components. Domain-relevant skills, creativity-relevant processes, and task motivation are individual components, meanwhile the social environment is the non-individual component. In [94], [95], [96] group creativity is seen as "a function of individual creative behavior "inputs", the interactions of the individuals involved (e.g. group composition), the group characteristics (e.g., norms, size, degree of cohesiveness), the group processes (e.g., approaches to problem solving), and the contextual influences (e.g. the larger organization, characteristics of group task)". Furthermore, the creative outcomes of groups and contextual influences affect organizational creativity. The features that contribute to group creativity are, at various levels: "individual (cognitive abilities/style, personality, intrinsic motivation, knowledge), group-related (cohesiveness, size, diversity, role, task, problem- solving approaches), and organizational (culture, structure, strategy, technology, resources, rewards etc.)".

Q-Learning algorithm for building creative students' groups

Description of the algorithm

The Q-learning algorithm introduced by Watkins is the most applied reinforcement algorithm with the help of which an intelligent agent learns an optimal action-selection policy [97]. The objective is to maximize the total reward through multiple iterations of interactions of the agent with the environment. The algorithm uses a Q-table containing pairs of state and action. It starts with an initial estimate $Q(s,a)$, s represents a state and a an action.

When a certain action a is selected in a state s, agent gets a reward $R(s,a)$ and the next state of the system is acknowledged.

The estimation of the function value-state-action is:

$Q(s,a) \leftarrow Q(s,a) + \alpha * \left(R(s,a) + \gamma * max_{a'} Q(s',a') - Q(s,a)\right)$

where $\alpha \in (0,1)$ is the learning rate and $\gamma \in (0,1)$ is the discount factor.

We introduced in [20], [22], [23] a Q-learning algorithm - based method to build in an optimal way the most creative learning groups (GC-Q-Learning), which is presented as follows (our approach used the basic form of Q-learning):

Notations

n – number of students

$c = (c_1, c_2, \dots c_m)$ - creativity vector, c_i represents a characteristic of students influencing group creativity, m – number of individual characteristics

id_group – the group identification

k – the number of groups

$(c_1, c_2, \dots c_m, id\_group)$ – a state (*s*), composed by the creativity vector and the group identification for each student

action (*a*) – the action of moving a student to another group in which he would contribute the most to increasing the group creativity

Q – expresses the quality of association between a state and an action, in the sense of our goal, to build the most creative k groups

R - reward is the value of group's creativity, and it ranges between 1 and 5.

Our goal was to obtain an organization of students in groups, in which either each group will have a creativity value larger than a desired threshold or the average creativity on all the groups will be higher than such a threshold.

The state space includes the set of tuples that can be built considering that each characteristic can have a finite number of values. The size of action space is given by the number of groups (k) to be constructed.

For this algorithm, a student is not a particular person, but a particular type of student given by her set of characteristics. Therefore, all the students having the same creativity vector will be a generic student for our algorithm.

1. Build a bi-dimensional matrix Q for all possible pairs $\langle state, action \rangle$: $(c_1, c_2, \dots c_m, id\_group, action\_number, q)$

   A value of the $action\_number$ equals with *i* means that if a particular type of student (given by their creativity vector $(c_1, c_2, \dots c_m)$) will be moved to the group having the value of id_group equal with *i*, then their contribution to group creativity is quantified by q (in this stage). All the elements in the q column may be initialized with 0 or with

a randomly chosen low value. On each line of the matrix, the data that corresponds to each type of student involved in the grouping process is included, i.e. the values of thier characteristics, the current group number, the action number, and the value computed for q *(that quantifies a potential for creativity)*. One particular type of student could have more corresponding lines, one for each combination $\langle current\ id\_group, action \rangle$

2. Initialize the *optimal_policy* with an initial policy. In our case, the optimal policy is an optimal grouping of students that maximizes group creativity. The initial grouping is set by the instructor and the students together and experience shows that they tend to group as cliques determined by their inter-personal affinities.
3. Group the students and have them carry out working sessions, in which each group's creativity is assessed, and its score is assigned to the reward $R(s,a)$ The values of $R(s,a)$ are obtained with help from human experts. The reward describes the potential of the groups' creativity. Then, the matrix Q is updated for each such working session. This procedure is presented below.

   procedure working_session_computation
   select action of (optimal_policy) /* student grouping*/
   compute $R(s,a)$
   compute table Q
4. Analyse the group creativity for each group against the global objective (the optimal grouping policy), which is getting closer to the maximum value possible for R, for each group or for all the groups. Re-iterate from step 3, if necessary.

Once the optimal policy consisting in tuples $(c_1, c_2, \dots c_m, id\_group)$ is obtained, an agent has learned to build the most creative groups.

Experiments and results

We performed more use cases with the algorithm GC-Q-Learning described above. In the first use case, we have undertaken a pedagogical experiment with 27 undergraduates and graduate students in Computer Science. The core of the experiment consisted in online brainstorming sessions discussing the curricula for our Computer Science programs, both at undergraduate and graduate level. We considered the motivation grade and creativity score

for individual characteristics assessed with specific questionnaires, and the group's creativity is evaluated by a teacher based on the ideas recorded in the brainstorming sessions with a score from 1 to 5. For the motivation grade, we used three classes: 0-low, 1-middle, 2-high and the creativity score ranged between -12 and 18. For the creativity score, we defined 5 classes: class 1 for score between -12 and -7, class 2 for score between -6 and -1, class 3 for score between 0 and 5, class 4 for score between 6 and 11, an class 5 for score between 12 and 18.

In later use cases, we extended the experiment, including three online brainstorming sessions with the subjects related to (1) the improvement of the curricula study programs, (2) the teaching and learning methods, and (3) the student lives on campus.

The values of the features for our pool of students are presented in Table 4.1.

Table 4.1. The values of the features describing students – creativity class and motivation level

Features (creativity class, motivation level)	Number of students
(2, 1)	6
(2, 2)	3
(3, 1)	9
(3, 2)	12
(4, 1)	6

We had 5 types of students with the following features: (creativity score=3, motivation level=1), (creativity score=3, motivation level=2), (creativity score=2, motivation level=1), (creativity score=2, motivation level=2), and (creativity score=4, motivation level=1) and investigated 9 possible groups described in Table 4.2.

Table 4.2. Groups' descriptions

id group	Student 1		Student 2		Student 3		Student 4	
	creativity class	motivation level	creativity class	motivation level	creativity class	motivation level	creativity class	motivation level
1	3	1	3	2	3	1	3	1
2	2	1	2	2	2	1	3	2
3	4	1	4	1	3	2	3	2
4	3	2	2	1	3	2	3	2
5	3	1	2	1	3	2	4	1
6	3	1	3	1	2	2	4	1

7	3	1	3	1	2	1	4	1
8	3	2	2	1	3	2	3	2
9	3	1	2	2	4	1	3	2

Group 1 was composed of 3 students with creativity class = 3 and motivation level = 1 and 1 students with creativity class = 3 and motivation level = 2;
Group 2 was composed of 2 students with creativity class=2 and motivation level =1 and 1 student with creativity class=2 and motivation level=2 and 1 student with creativity class 3 and motivation level =1, and so on.
The results of the computed Q-values (after 3 sessions of grouping) are shown in Table 4.3.

Table 4.3. Final computed Q-values

creativity class	motivation level	id of the group	Q value
3	1	1	3.46875
3	1	2	0
3	1	3	0
3	1	4	0
3	1	5	2.6971875
3	1	6	3.29578125
3	**1**	**7**	**3.79828125**
3	1	8	0
3	1	9	2.5321875
3	2	1	1.5
3	2	2	2.375
3	2	3	4.2340625
3	2	4	4.477402344
3	2	5	2.888515625
3	2	6	0
3	2	7	0
3	**2**	**8**	**4.872613525**
3	2	9	2.883153381
2	1	1	0
2	**1**	**2**	**3.5**
2	1	3	0

2	1	4	1.9
2	1	5	2.705
2	1	6	0
2	1	7	2.54
2	1	8	2.54
2	1	9	0
2	2	1	0
2	2	2	2
2	2	3	0
2	2	4	0
2	2	5	0
2	2	6	1.83
2	2	7	0
2	2	8	0
2	**2**	**9**	**2.165**
4	1	1	0
4	1	2	0
4	**1**	**3**	**3.78875**
4	1	4	0
4	1	5	2.7771875
4	1	6	2.2771875
4	1	7	2.6121875
4	1	8	0
4	1	9	2.6121875

The interpretations of Q-values are:

A student with the features (creativity class=3, motivation level=1) would contribute the most to the group creativity if they were in group 7, and decreasingly - in group 6, 1, 5 or 9.

A student with the features (creativity class=3, motivation level=2) would contribute the most to the group creativity if they were in group 8, and decreasingly - in group 4, 3, 5, 9, 2 or 1.

A student with the features (creativity class=2,motivation level=1) would contribute the most to the group creativity if they were in group 2, and decreasingly - in group 5, 7, 8 or 4.

A student with the features (creativity class=2, motivation level=2) would contribute the most to the group creativity if they were in group 9 and decreasingly - in group 2 or 6.
A student with the features (creativity class=4, motivation level=1) would contribute the most to the group creativity if they were in group 3, and decreasingly - in group 5, 7, 9, or 6.
The optimal building of the groups with regard to their creativity depends on the context of the learning process. The results show that adapted Q-learning algorithm offer a solution to generate the most creative groups of students in a specific situation. The performed experiments are detailed in [20], [22], and [23].

A multiagent system for building creative students' groups

The GC-Q-Learning presented above was integrated into a multiagent system, namely GC-MAS [20], [23]. Here, we briefly present the architecture of GC-MAS and the main roles of the 5 agents that compose it. The goal of the GC-MAS is to group students in the most creative teams. The agents composing GC-MAS are:

- The Communication Agent (CommGC) has a dual role, being responsible with interfacing with the users (both students and instructors) and with the agents, along with managing the activities of the other agents;
- The Creative Groups' Builder (BuildGC) is an agent that assists the construction of creative groups based on a reinforcement learning algorithm;
- The Creativity Evaluation Agent (EvalGC) assesses each group creativity;
- The Creativity Booster (EnvrGC) boosts development and maintenance of contextual environments that provide for increasing group creativity;
- The Facilitator Agent (FcIGC) facilitates a more efficient group interaction, e.g. by sustaining the team members who are shyer or less active. It also provides support for seeking out and taking on otherwise neglected tasks that have potential to facilitate creative group performances.

All agents are task agents, except the CommGC, which acts as a middle agent. The overall architecture is presented in Figure 4.1.

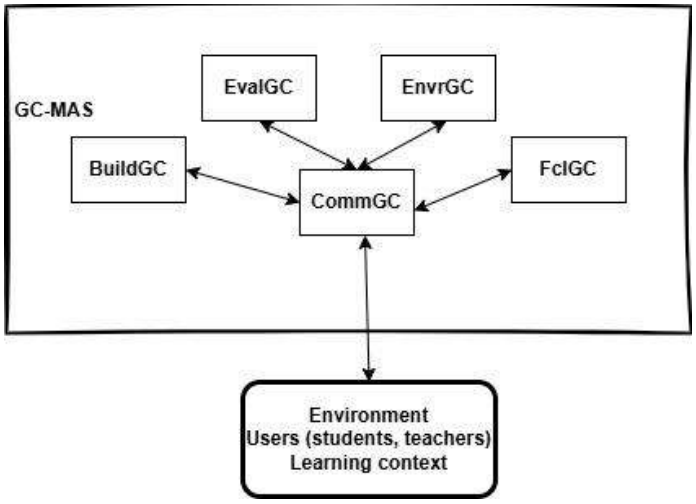

Figure 4.1. GC-MAS - the bird's eye view architecture [23]

The agents BuildGC, EvalGC, EnvrGC and FclGC are execution agents that perform precise actions in the construction of creative groups. They have a very simple structure, are goal-oriented, and use plan libraries or other machine learning algorithms to perform their roles.

CommGC has a horizontally stratified structure, in which each level is connected directly to both the input sensors and the output effectors. Each level acts as an individual agent that provides the expected action. The two levels of CommGC are as follows: (1) the social level that ensures communication with the other agents, the users, and with the external environment, as a true personal/interface agent, and (2) the administrative level that coordinates the actions of all the agents.

Bayes classifiers for building creative students' groups

In this section we present a model and a method to group students in the best teams in collaborative learning situations using Bayes classifications [21]. We applied them in a collaborative learning context considering that the best teams refer to the most creative students' teams. The method proposed could be extended to any given collaborative work situation.

To develop the model for building innovative groups, we analysed more approaches, of which we reported as the most relevant for our problem the references [98], [99], [100]. We considered four attributes to obtain the most creative and innovative students' teams. Three of them were related to the individual (creativity, motivation and domain knowledge) and the last one is related to the inter-personal affinities.

The main idea of our approach was to consider a group characteristic that was the most relevant for the proposed goal and to maximize it by repeatedly grouping people based on

the values of some individual features. Our model comprised 3 stages which are presented in Figure 4.2.

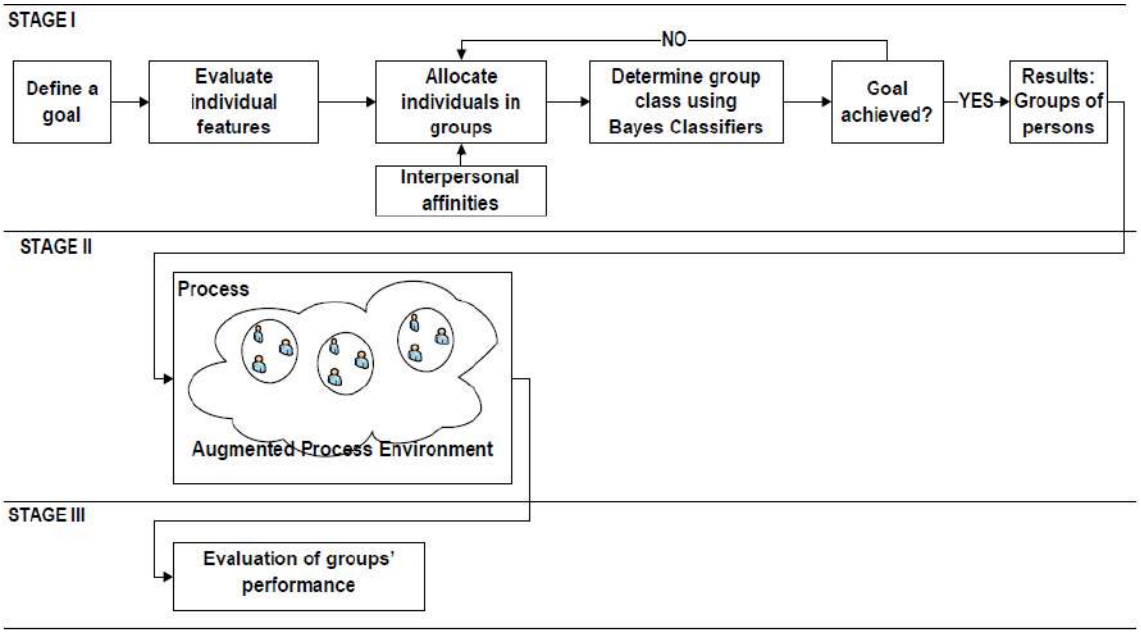

Figure 4.2. Model for building –the best collaborative groups [21]

- In the Stage I, each learner was allocated to a group of students. For each group we made predictions using Bayes classifiers. Two cases could appear. If the classes of the groups satisfied the proposed goal (for example, given 5 groups, 2 of them were high level of creativity, 2 were medium creativity and 1 was a low class of creativity) then the distribution stopped. If the proposed goal was not satisfied, then we tried another combination of student groups.
- In the Stage II, we increased the groups' creativity by augmenting the contextual environment through specific tasks, different instructional strategies, questions and answer sessions, and so on.
- Finally, in Stage III, the performance of each group was assessed using specific measurement. If the goal was achieved, then the most creative teams were built, otherwise the groups would be re-organized in the next instructional session.

To build an instance for our model, we assessed the learners' attributes with impact of group creativity: individual level of creativity, personal motivation, domain expertise (Figure 4.3).

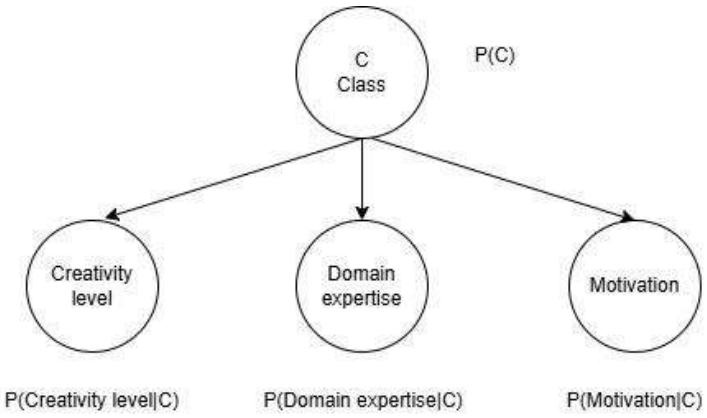

Figure 4.3. The Naïve Bayes structure for groups classification [21]

We have tested the proposed model in a real-world scenario: our goal was to group 20 students of CS study program in the best creative teams. The values of students' features used in this case are presented in Table 4.4.

Table 4.4. Sophomores' attributes used in classification [21]

Learner Id	Creativity score	Domain Expertise	Motivation	Learner Id	Creativity score	Domain Expertise	Motivation
Learner 1	5	8	2	Learner 11	8	10	1
Learner 2	4	8	1	Learner 12	-2	7	1
Learner 3	7	8	2	Learner 13	-1	6	2
Learner 4	7	10	2	Learner 14	7	7	1
Learner 5	8	8	1	Learner 15	4	8	1
Learner 6	3	8	2	Learner 16	0	5	2
Learner 7	2	7	0	Learner 17	5	5	2
Learner 8	2	6	0	Learner 18	3	5	0
Learner 9	2	6	1	Learner 19	-5	6	0
Learner 10	-2	5	0	Learner 20	-6	6	0

The creativity score was evaluated using Gough's Creative Personality Scale [128], which output range is between -12 and 18, and the motivation using an adapted questionnaire based on MSLQ [129]. The values for domain expertise were the grades obtained at the Data Structures and Algorithms course, while the students were freshmen.

We used two Bayesian classifiers: one to predict creativity class for each student and other to predict the creativity class for each group. We considered 3 classes for level of creativity codded as 1 – low level of creativity, 2 – medium level of creativity and 3 high level of

creativity. Our objective was to obtain at least three teams with creativity class medium or higher.

Starting from data presented in Table 4.4, we re-grouped students using the two Bayes classifiers until our objective was achieved.

Finally, we obtained three groups of students from four with level of creativity medium or high:

- group 1 composed of learners 1, 2, 4, 9, 18.
- group 2 composed of learners 11, 12, 14, 15, 19.
- group 3 composed of learners 3, 5, 7, 8, 13.
- and one group (group 4) composed of learners 6, 10, 16, 17, 20 with a low level of creativity (Figure 4.4).

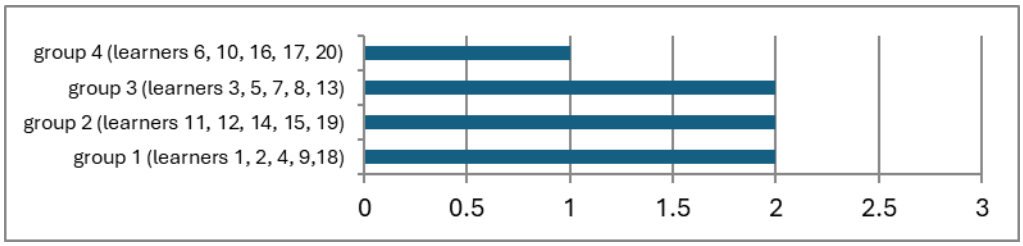

Figure 4.4. Creativity classes for the groups [21]

Moreover, to validate our model we performed an evaluation of the learning achievements before and after the grouping and we observed that the performance of the groups was higher after this collaborative learning experience (Table 4.5).

Table 4.5. Evaluation of learning achievements before and after the grouping [21]

Group	Before grouping	After grouping
Group 1	7.40	8.20
Group 2	7.60	9.40
Group 3	7.00	8.40
Group 4	5.80	8.00

The results that can be obtained with our model depend on the values of the students' attributes, the learning context, the organizational factors, and the students and teachers' emotions. In the next chapter we consider these emotional factors, and we develop AER for the emotions often encountered in the learning process.

Chapter 5 Automated Emotion Recognition in the Learning Process

This chapter presents the results of the research started in 2023 in the project entitled Affective Learning: Benefits and Ethical Risks in Superior Education (ALBER) conducted by the author of this thesis. The usage of automatic emotion recognition in the educational process may be beneficial to students; research and literature in the field highlighting and validating the benefits of their integration into the learning systems. The risks of using AER in the educational process are a less studied topic, although there are numerous social, moral or ethical implications when integrating AER in educational systems. The ALBER project focuses on the investigations of the affective learning field, trying to draw a leverage between the benefits and ethical risks of using AERs in the learning process. The project was carried out within the Computer Science&Education Laboratory, Petroleum-Gas University of Ploiești. More studies were conducted, and more experiments were performed related to the usage of AER in learning in compliance with the rules adopted by European Commission in AI Act (2024). A part of the findings of our research related to the usage of AERs in educational systems are published in [24], [25].

Findings

- 6 categories of AER roles in education.
- 1D-CNN model using 5 EEG channels from DEAP dataset to recognition emotions often encountered in education. The accuracies for the model are boredom – 99.64%, confusion – 99.69%, frustration – 99.65%, curiosity - 99.79%, excitement – 99.90%, concentration – 99.69%, anxiety – 99.20%. The running time is under 9342.58 seconds for all emotions' cases. The F1 score varies from 91.76%, in the case of anxiety to 99.07%, in the case of concentration.
- ethical model for AER in online learning.

The Importance of the topic

Undoubtedly, there is a complex relationship between emotion and the learning process as more studies demonstrate [26], [47], [101], [102], [103], [104]. A wide range of emotions are engaged in an academic context: anxiety, hopefulness, hopelessness, relief, enjoyment of learning, pride of success, anger, shame, boredom, surprise, sadness, frustration, confusion,

happiness, fear, joy, disgust, interest, curiosity, contempt, delight and excitement [26], [27], [28]. Even though in most cases, positive emotions as happiness have a beneficial effect on learning and negative ones as frustration have the opposite effect, this is not a rule [105]. For example, negative emotions, as confusion, can bring benefits in the learning process if they are managed correctly [106].

The affective learning considers the implications of emotions in the learning process through affect-aware technologies. Yadegaridehkordi et al. (2019) perform a review of the literature on AC in education and select 94 studies (published in the period 2010-2017) according to the keywords "affective computing in education/learning, emotion in education/learning, students'emotion, emotion recognition" to find the answers for four research questions: "(RQ1) What are the trends in affective computing in education/learning as can be gauged from the selected papers? (RQ2) What are the main research purposes and learning domains addressed the selected papers? (RQ3) What are the main affective measurement channels and methods used in the selected papers? (RQ4) What are the major theories/models of emotion adopted and the emotional states considered in the selected papers?" [27]. The papers were selected from ISI Web of Knowledge, ScienceDirect, IEEE Explorer, and Springer Link databases. Their findings show:

- RQ1 – an increasing interest for the research topics related to AC in education.
- RQ2 – the subjects related to the emotion recognition and expression systems and relationship between emotion, motivation, learning style and cognition are in the spotlight.
- RQ3 – the channels used to emotion assessment, in order of usage, are textual (ER using self-reporting questionnaires or text or expert observation – 50 studies), multimodal (any combination of others channels – 16 studies), physiological (ER from EEG, ECG, HRV, SCL signals, and eye tracking – 13 studies), visual (ER from facial expression – 8 studies), and vocal (ER from speech and intonation – no study).
- RQ4 – the dimension models of emotion are highly used and the ordered emotions in terms of frequency analysis emotions are: boredom – 41; anger – 37; anxiety – 30; enjoyment – 27; surprised – 25; sadness, frustration – 24; pride - 21; hopefulness – 20, hopelessness, shame – 19; confusion, happiness, natural, fear, joy – 17; disgust, interest – 15; relief, excitement – 13.

Mello&Graesser (2015) highlight the importance of the automated affect-detection and response technologies to regulate both positive and negative emotion [3]. Called the reactive

affect-aware technologies, some examples include Affective AutoTutor, GazeTutor, and Miters.

Affective AutoTutor is an Intelligent Tutoring System (ITS) which helps students master difficult concepts in Newtonian physics through "a human-like interactivity" based on "automatically detecting and responding to students' emotional states in addition to their cognitive states" [107].

GazeTutor is also an ITS, which uses an eye tracker to detect learners' emotions, bored, disengaged, or zoning out, and react appropriately [108].

Miters – "Multimodal Intelligent Tutoring Emotion Recognition System" is an affect-aware ITS integrated in e-learning environment named "Smart Learning Room" [109]. Miters use three channels, face, text, and speech, to recognize four emotions neutral, joy, sadness, and anger. The teachers are noted by the students' emotions and react to handling the learning situation.

Even though there is a clear increasing trend towards introducing affect-aware technologies into education, the ethical impact of this is poorly discussed and researched. The most research related to affective learning addresses the development of performant AER and the benefits for the process, and the ethical implications are generally neglected.

So, we claim that machine ethics is an essential aspect needed to be considered in affective computing (Figure 5.1.) [24].

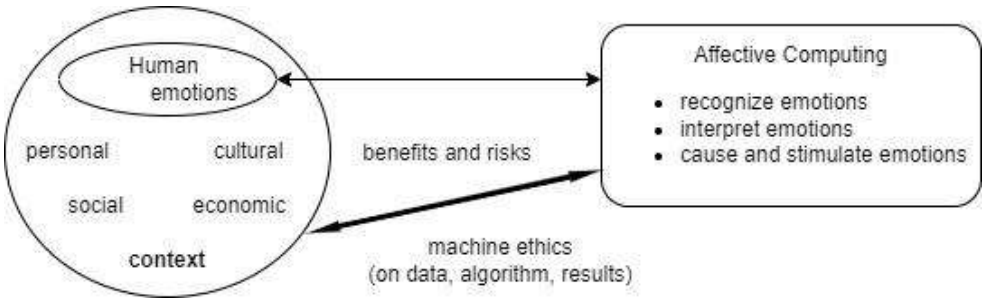

Figure 5.1. Ethics in affective computing [24]

Leveraging the benefits and risks of integrating affect-aware technologies in education requires performant AERs with at least four attributes: reliable, non-discriminatory, and less intrusive [24]. Furthermore, there are necessary to provide the explanations for the predictions of ML models.

Roles of AER in education

We identified in the [24], six categories of roles of AER systems in education (Table 5.1). To achieve an understanding of the ways in which AER systems are used in education, we performed a search with key phrase "emotions recognition" AND "online learning" in ScienceDirect for the period 2015-2022. 102 papers were returned, after abstract and full text screening, we found 27 studies relevant for our purpose (26 studies dealing with students' emotions and only one with teachers' emotions).

Table 5.1. AER roles in online learning [24]

AER roles
Development of Intelligent Tutoring Systems (ITS) with emotional abilities, able to detect emotions and react appropriately (the relationships between students' emotions, motivation, cognition, learning styles are speculated)
Supporting engagement and motivation of learners and teachers (both students and teachers need to be aware of their emotional and mental states)
Learning assessment (detect cheating by students)
Teaching assessment (teachers need to be aware of the impact of emotions on the teaching-learning process)
Building comfortable learning environments (increasing the efficiency of learning and teaching)
Supporting students with special needs (ADHD, anxiety and so on)

We found several experiments described in literature to highlight the benefits of using AER in education. Some of them presented in [24] are:

- "Learning Companion" project, Affective Computing Group, MIT Media Lab - a "computerized Learning Companion" model to react appropriately at the affective state of learners [110].
- An experimental study carried out in a Shanghai online college, the learners' emotions are recognized based on physiological signals [111].
- Afective AutoTutor – controlled experiments prove improvements in the learning process [107].
- Students' engagement maintaining using deep learning models based on facial expressions [112].

- Students' engagement recognition in a controlled experiment (math learning) [113].

1D-CNN model for emotion recognition based on 5 EEG channels - DEAP dataset

We obtained a highly accurate model for the seven emotions (boredom, confusion, frustration, curiosity, excitement, concentration, and anxiety) recognition using only 5 EEG channels from DEAP dataset [25].

To achieve it, we performed an investigation of the most performant AER models using EEG values from DEAP and found the model proposed by Akter et al. (2022), a 1D-CNN model with four convolutional layers and three dense layers using 14 EEG channels (FP1, FP2, AF3, Fz, F3, F4, F7, F8, FC1, C4, P3, P4, PO3, PO4) [113]. The accuracies for model are for valence - 99.89%, and for arousal – 99.83%. We used Akter et al. (2022) model as basis for our approach.

For describing emotions we used PAD dimensions according to [61] (chapter 2) and made a rough estimation for the minimum and maximum of PAD values from DEAP (Table 5.2). We consider this aspect a limitation of our study because the data has not been acquired and labelled within a learning scenario.

Table 5.2. Minimum and maximum PAD values in DEAP range for the seven emotions [25]

Emotion	Pleasure		Arousal		Dominance	
	min	max	min	max	min	max
Boredom	1.64	3.16	1.56	3.48	2.84	4.52
Confusion	2.08	3.68	4.92	7.24	2.6	4.84
Frustration	1.72	3.16	5.6	8.56	2.4	4.8
Curiosity	4.68	7.08	6.68	8.28	3.6	6.32
Excitement	6.48	8.48	7.2	8.8	5.36	7.68
Concentration	5.68	7.68	5.04	7.2	5.32	7.8
Anxiety	3.24	6.84	6.12	8.6	3.12	5.68

Using the values from above, we labelled DEAP records with 1 or 0, depending on the presence or non-presence of one of the following emotions: boredom, confusion, frustration, curiosity, excitement, concentration, anxiety.

The stages of development AER using EEG signals are presented in Figure 5.2.

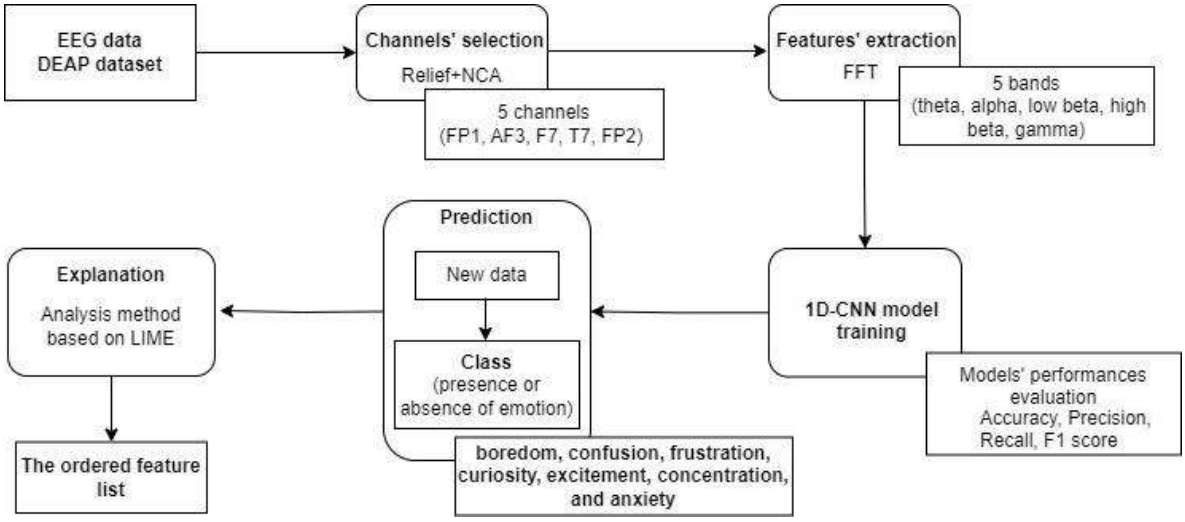

Figure 5.2. Our framework for AER in the learning process [25]

EEG channels selection

To reduce the complexity of the inputs, meaning to use an as small as possible number of EEG channels from DEAP, we considered the top 10 channels relevant for emotion recognition found by Topic et al. (2022) using ReliefF and NCA (Neighborhood Component Analysis) methods [114].

Table 5.3. Top 10 channels for emotion recognition [114]

ReliefF	**FP1**	**AF3**	F3	**F7**	**T7**	O1	Oz	**FP2**	F8	P8
NCA	**FP1**	**AF3**	**F7**	**T7**	CP5	P7	**FP2**	AF4	FC6	T8

We proceeded to look at the intersection of the two EEG channels sets and selected the following channels: FP1, AF3, F7, T7, FP2 (Figure 5.3).

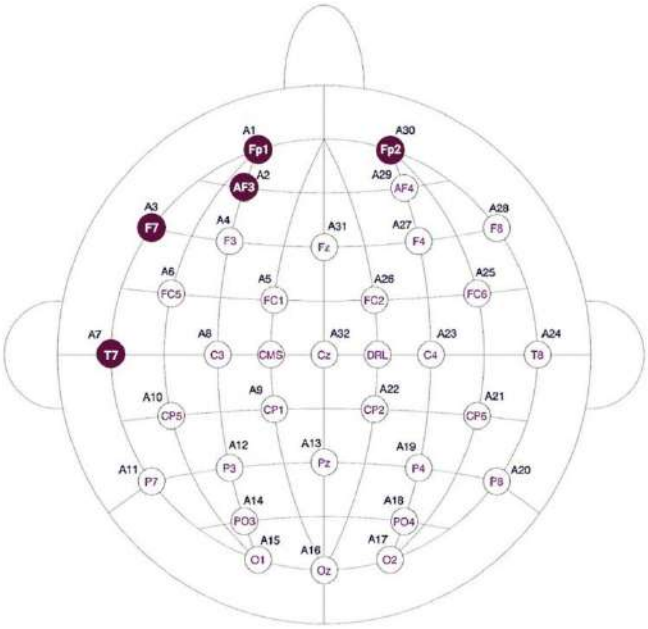

Figure 5.3. The 5 EEG channels selected for our algorithm [25]

Features extraction and processing

For features extraction, we used FFT (Fast Fourier Transform) as in [113] with channels FP1, AF3, F7, T7, FP2 EEG channels and sub-bands theta 4-8Hz; alpha 8-13Hz; low beta 13-22Hz; high beta 22-30Hz, gamma 30-100Hz. To perform this, we used PyEEG library, function bin_power with sampling rate equals with 128 Hz. The features' values were standardized by removing the mean and scaling to unit variance. A capture from data is presented in Figure 5.4.

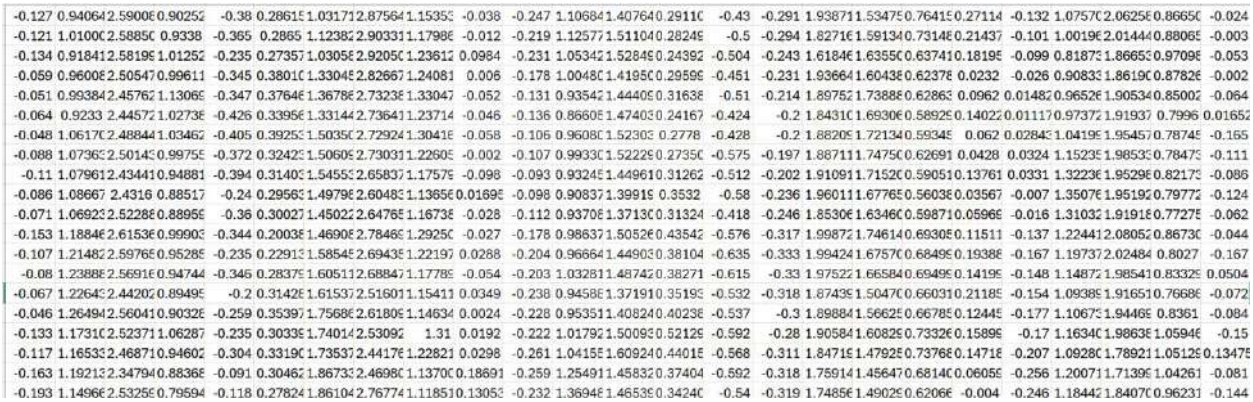

Figure 5.4. A capture from processed data – 5 EEG channels * 5 sub-bands

The selection and training of the ML model

We adapted the 1D-CNN model from [113] for 5 EEG channels. The number of chosen features was 25 (i.e. 5 EEG channels x 5 sub-bands). For the 4 convolutional and 3 dense layers, the activation function is ReLU and for the output layer is Softmax. For training, we set batch_size = 100, arbitrary epochs = 100 and implemented early stopping techniques to reduce the training time. Using early stopping techniques, we avoided overfitting. The training of the 5 channels models stopped after 42 epochs for anxiety, 57 epochs for concentration, 34 for confusion, 55 for curiosity, 34 for excitement, 29 for frustration, and 51 for boredom. Our model is described in Table 5.4.

Table 5.4. The 1D-CNN model used for seven emotions recognition [25]

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 25, 1)]	0	[]
conv1d (Conv1D) filters=32, kernel_size=6	(None, 25, 32)	224	['input_1[0][0]']
batch_normalization	(None, 25, 32)	128	['conv1d[0][0]']

max_pooling1d pool_size=2	(None, 12, 32)	0	['batch_normalization[0][0]']
conv1d_1 (Conv1 D) filters=32, kernel_size=6	(None, 12, 32)	6176	['max_pooling1d[0][0]']
batch_normalization_1	(None, 12, 32)	128	['conv1d_1[0][0]']
max_pooling1d_1 pool_size=2	(None, 6, 32)	0	['batch_normalization_1[0][0] ']
conv1d_2 (Conv1D) filters=32, kernel_size=6	(None, 6, 32)	6176	['max_pooling1d_1[0][0]']
batch_normalization_2	(None, 6, 32)	128	['conv1d_2[0][0]']
max_pooling1d_2 pool_size=2	(None, 3, 32)	0	['batch_normalization_2[0][0] ']
concatenate	(None, 21, 32)	0	['max_pooling1d[0][0]', 'max_pooling1d_1[0][0]', 'max_pooling1d_2[0][0]']
conv1d_3 (Conv1D) filters=128, kernel_size=6	(None, 21, 128)	24704	['concatenate[0][0]']
batch_normalization_3	(None, 21, 128)	512	['conv1d_3[0][0]']
max_pooling1d_3 pool_size=2	(None, 10, 128)	0	['batch_normalization_3[0][0] ']
flatten (Flatten)	(None, 1280)	0	['max_pooling1d_3[0][0]']
dense (Dense) units=1024, activation='relu'	(None, 1024)	1311744	['flatten[0][0]']
dropout (Dropout) rate=0.2	(None, 1024)	0	['dense[0][0]']
dense_1 (Dense) units=256, activation='relu'	(None, 256)	262400	['dropout[0][0]']
dropout_1 (Dropout) rate=0.2	(None, 256)	0	['dense_1[0][0]']
dense_2 (Dense) units=64, activation='relu'	(None, 64)	16448	['dropout_1[0][0]']
dropout_2 (Dropout) rate=0.2	(None, 64)	0	['dense_2[0][0]']
dense_3 (Dense) units=64, activation='softmax '	(None, 2)	130	['dropout_2[0][0]']

Trainable params: 1,628,450

Non-trainable params: 448

Prediction

We used the model to predict the presence or absence of one of the seven emotions for various instances of data.

Explanation of the predictions – our analysis method based on LIME

We proposed a LIME-based method to explain the outcomes for the 1D-CNN model. The method and the explanations are presented in Chapter 6, dedicated to interpretability and explainability techniques.

The performance of 1D-CNN model

Table 5.5 presents the performance of the 1D-CNN model based on 5 EEG channels, namely FP1, AF3, F7, T7, FP2. The running time was under 9342.58 seconds for all emotions' cases, and there can be noticed that all the test accuracy values are over 99.21%, The F1 score varies from 91.76%, in the case of anxiety to 99.07%, in the case of concentration.

Table 5.5. The performance of the 1D-CNN model (5 EEG channels) for the recognition of the seven emotions [25]

Emotion	Performance (%)					Running time (seconds)
	Test loss	Test accuracy	Precision	Recall	F1_score	
Boredom	0.91	**99.64**	94.62	97.41	**95.97**	8281.34
Confusion	1.16	**99.70**	99.17	96.19	**97.63**	6004.65
Frustration	0.98	**99.66**	98.98	95.13	**96.97**	5690.35
Curiosity	0.97	**99.80**	99.74	96.83	**98.24**	8675.36
Excitement	0.41	**99.91**	98.74	97.75	**98.24**	5790.36
Concentration	1.14	**99.70**	99.16	98.98	**99.07**	9342.58
Anxiety	1.91	**99.21**	95.03	88.96	**91.76**	6655.65

The confusion matrices for the seven emotions' recognition (5 channels + 1D-CNN model) are presented in Figure 5.5.

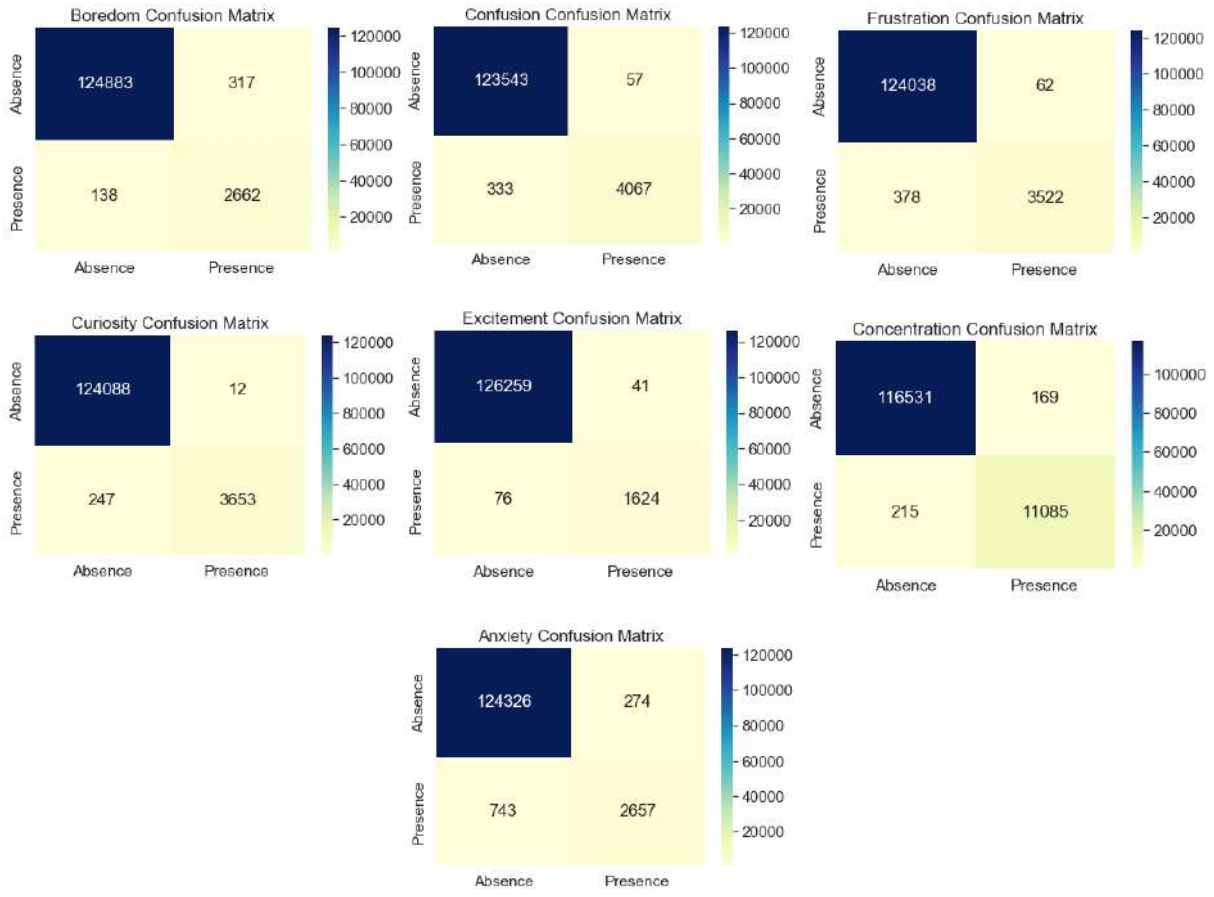

Figure 5.5. The confusion matrices for the seven emotions (5 channels + 1D-CNN model) [25]

Our conclusion is: using only 5 EEG channels from DEAP dataset is sufficient to obtain high performance trained models for emotion recognition.

We performed more comparisons between 1D-CNN model developed on 5 channels (our achievement), 14 channels as in [113] as well as 10 channels selected with ReliefF technique, and 10 channels selected with NCA as in [114] (Table 5.6., 5.7. for model performance, Table 8 for running time).

Table 5.6. The accuracy for 1D-CNN models [25]

	boredom	confusion	frustration	curiosity	excitement	concentration	anxiety
14 channels [113]	99.91	99.87	99.92	99.87	99.98	99.87	99.93
10 channels-ReliefF [114]	99.81	99.87	99.69	99.85	99.74	99.81	99.74
10 channels-NCA [114]	99.91	99.60	99.84	99.69	99.92	99.69	99.64
5 channels	**99.64**	**99.70**	**99.66**	**99.80**	**99.91**	**99.70**	**99.21**

Table 5.7. F1 scores for 1D-CNN models

	boredom	confusion	frustration	curiosity	excitement	concentration	anxiety
14 channels [113]	99.36	98.80	91.89	99.65	99.42	99.77	99.52
10 channels-ReliefF [114]	97.81	99.03	97.41	98.77	94.79	99.42	97.53
10 channels-NCA [114]	98.94	96.92	98.61	98.57	98.41	99.05	99.40
5 channels [25]	**95.97**	**97.63**	**96.97**	**98.24**	**98.24**	**99.07**	**91.76**

The running times for training of the models decreases considerably with one exception (curiosity), as one can notice in Table 5.8.

Table 5.8. Running time for training of the models (seconds)

	boredom	confusion	frustration	curiosity	excitement	concentration	anxiety
10 channels-ReliefF [114]	12487.59	10117.58	8421.99	10196.85	9333.75	16821.48	13403.73
10 channels-NCA [114]	14119.25	10057.39	12463.57	7011.58	7638.51	12405.46	9943.37
5 channels [25]	**8281.34**	**6004.65**	**5690.35**	**8675.36**	**5790.36**	**9342.58**	**6655.65**

The performance achieved by 1D-CNN model with 5 channels is comparable with the models based on 14 and 10 channels. The running time for training the models decreases considerable, minimum about 24% and maximum about 54%.

Ethical model for AER in online learning

We proposed in [24] an ethical model for AER in online learning (Figure 5.5). The model was designed considering the 16 classes of ethical risks associated with the usage of AER in the learning process. The risks classes, detailed in [24], are:

- "Bias and discrimination,
- Unreliable, unsure, unsafe or poor results.
- Non-transparent, unexplainable, unjustifiable or not fully predictable outcomes.
- Privacy invasion by (1) inaccurate ownership and management of personal data, (2) failure in giving and withdrawing consent and (3) domestic surveillance.
- Unfairness and digital division.
- Deception.
- Manipulation and building authoritarian relations.
- Changes in human perception of reality, understanding, expertise and natural behaviour.
- Erroneous portraying of human beings and emotions.
- Denial or bypassing of individual autonomy and rights (restriction on users' ability to exercise free will or free speech, non-free and non-informed decisions regarding users, denial of right against self-incrimination).
- Dual use.
- Isolation of individuals, disintegration of social connections and dehumanizing of people relations by emotional and social interaction with high-performance, yet lacking self-awareness, AI systems.
- Dependence on a machine.
- Risk of losing the sense of individual identity.
- Replacement of the teachers.
- Lack of energetic sustainability".

The model uses the ethical framework develop by Leslie (2019), representing a "real tenet for the responsible AI systems" [29]. The model is based on Leslie's guide, the Ethics sheet on the sentiment analysis carried out by Mohammad [115], and the data ethics framework of the Central Digital and Data Office [116]. In the development chain of AER systems, humans must be truly ethical, as obtaining products with minimal ethical risks depends on them.

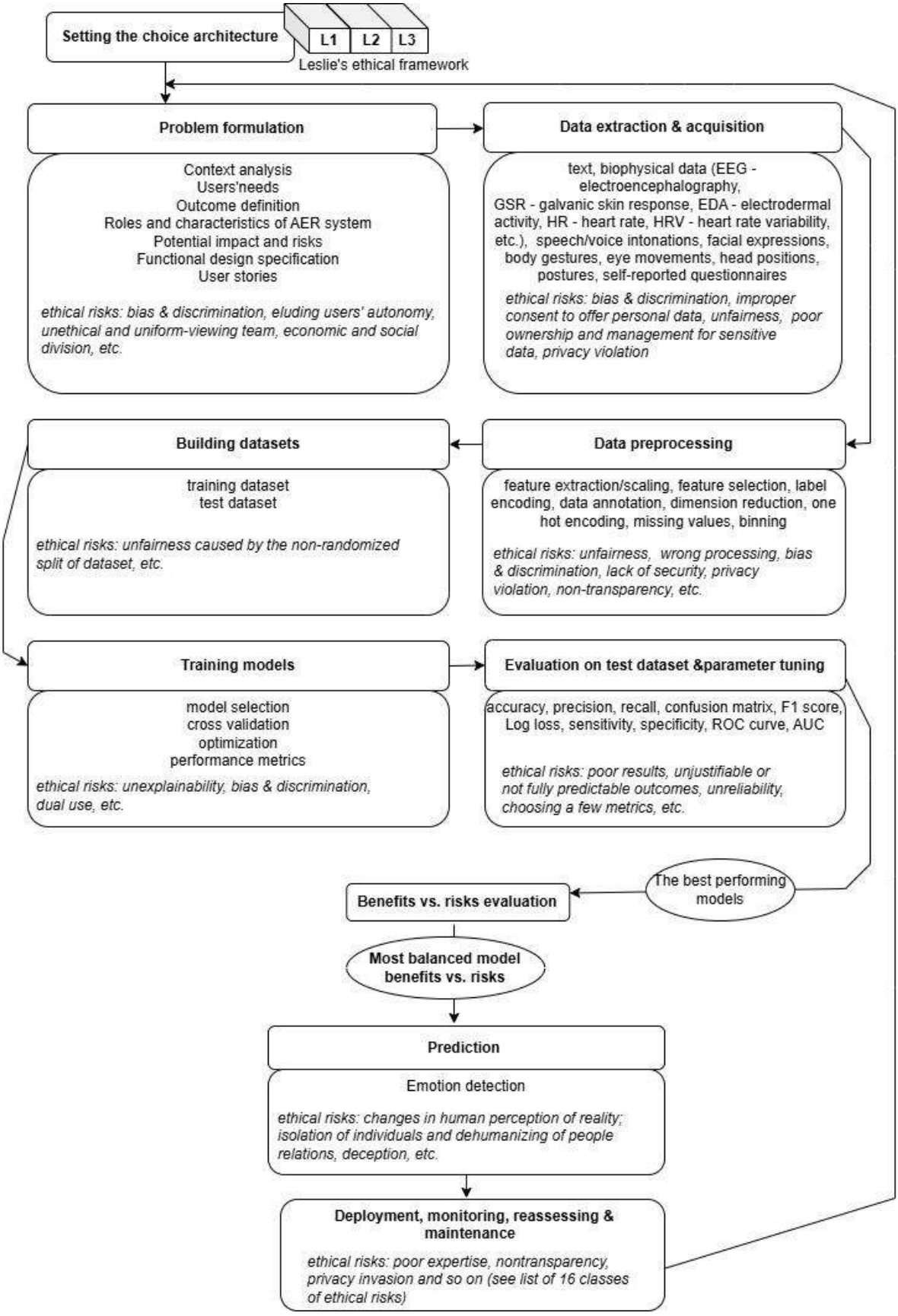

Figure 5.5. Ethical model for AER in online learning [24]

The Leslie (2019) ethical framework for AI comprises three levels [29]. The first level (L1) is grounded in ethical values that "Support, Underwrite, and Motivate Values – SUM Values" the responsibility emphasizing the impact of the AI on the society. The second level (L2) considers "Fairness, Accountability, Sustainability, and Transparency – FAST Track Principles" to develop responsible AI technologies. The third level (L3), called "Process-Based Governance - PBG Framework" offers practical ways to incorporate SUM Values and Fast Track Principles in development of the AI-based projects.

We must mention that the first and most important thing is the need for people involved in the development of AI projects to have ethical behaviour.

Our model follows the ML model creation pipeline with the additional consideration of ethics and the three levels (L1, L2, L3) at each stage to mitigate possible ethical risks. Project monitoring during both its development and implementation is a necessity to avoid the issues related to concept and model drift and possible ethical risk situations.

Chapter 6 Explainability in Machine Learning

This chapter presents the results of research related to explainability in machine learning published in [30] and [25]. It is a hot topic especially in the context of ensuring ethics and European regulation General Data Protection Regulation - GDPR (2016) [117] and AI Act (2024) [4]. We state with all confidence that humans have the right to know the reasoning of AI decisions, which affect them and their communities. Also, for AI practitioners knowing the reasoning behind the outcome of an AI decision offers a coherent understanding of the behavior of the AI model.

Findings

- A LIME-based analysis method to provide explanations for the 1D-CNN model's predictions presented in chapter 5.
- A cluster-based aggregation method for explainers. The code and results are publicly available at https://github.com/oanabalan/AggregateExplanations.

The Importance of the topic

AI becomes ubiquitous in everyday life. For most people, the outputs of AI systems are mysteries, capable of causing distrust and confusion. To get insight into AI algorithms and to understand the mechanism behind decisions is a difficult task, even for AI practitioners. Explainable Artificial Intelligence (XAI) tries to solve this issue and make AI algorithms more transparent. More complex ML algorithms assure high performance, but they are more opaque. XML (eXplainable Machine Learning) aims to leverage between performance and explainable models. "Interpretability" and "explainability" are often used interchangeably, even they are distinct. A well adopted definition for interpretability of a model is provided by Miller (2019) as "the degree to which an observer can understand the cause of a decision" [118]. Biran&Cotton (2017) consider the ability of an AI systems to "explain" its outputs as a key component and an AI system is "interpretable if their operations cand be understood by a human, either through introspection or through a produced explanation" [119]. In [120], an interpretation is seen as a correspondence between an output, the abstract concept, into an understandable form for humans and an explanation represented the set of features which contributed for a specific instance to model's decision-making.

The necessity for XAI is well covered by Adadi&Berrada in [121] considering four directions: explain to justify the decisions, explain to control and correct possible errors, explain to improve the models, explain to discover new knowledge. The demands for interpretations and explanations are required by human curiosity and learning; scientific curiosity; bias detecting; safety assurance; social acceptance; privacy, trust, and reliability ensuring; debugging and auditing; social acceptance and social interaction management [122].

Besides the fact that everyone wants to know why, there are the legal compliances, and we must note the European GDPR (2016) and AI Act (2024). Juliussen (2025) makes an investigation over the concept of the right to an explanation under GDPR and AI Act and concludes that:

"To recapitulate, both Articles 13, 14, and 15 contain a right to receive meaningful information about the logic involved when processing personal data as part of automated decision-making under Article 22 of the GDPR. However, Articles 13 and 14 of the GDPR apply to the collection of personal data, while Article 15 contains reactive rights that are dependent on requests from the data subject. Hence, Articles 13 and 14 on the one hand and Article 15, on the other hand, will provide different types of explanations where the first is more general ex-ante AI system descriptions and the latter is ex-post explanations of the output. Article 15 requires the controller to provide "meaningful" information. This condition entails that the information must be understandable, useful, reliable, and helpful for assessing the lawfulness of the data processing for the subject. It must be evaluated contextually on a case-by-case basis" [123]. "To conclude, the right to an explanation under Article 86 (1) of the AI Act covers outputs from high-risk AI systems in Annex III when the output forms the basis for a decision affecting natural persons" [123].

The FAT (fairness, accountability, and transparency) conceptual model is required to be considered and operationalized both in industry and academia [31]. The "algorithmic affordance" allows users to accept the results of the algorithm. The FAT conceptual model comprises critical factors, among which we also find understandability, explainability, and observability (Figure 6.1).

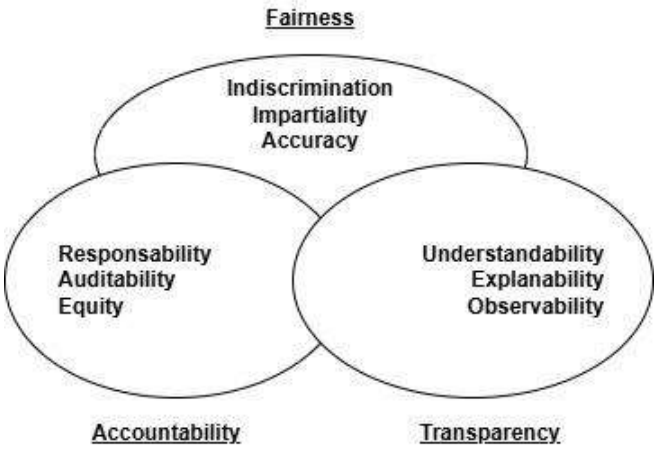

Figure 6.1. FAT model [31]

In AI ethics, transparency is the most requested attribute to obtain the evidence related to fairness of the algorithms [124]. The responsible AI paradigm, along with AI principles, focuses on ensuring that FAT factors are employed.

The landscape of XAI methods is dynamic. Usually, the most accurate ML models are not the most interpretable models, for example, deep neural networks, random forest vs. linear or logistic regression. Related to XAI methods, an issue related to the reliability of these methods needs to be solved. The outputs of various explainers applied at the same instance are different. Furthermore, the same explainer running on the same instance generates different results. The disagreement problem in ML consists in obtaining contradictory explanations for the same sample and model [32], [33], [34]. There is a challenge to obtain good explanations for the results of ML models, most practitioners apply various methods and perform an agreement between the outputs.

A LIME-based analyse method for explanations

Ribeiro et al. (2016) proposed an algorithm that explains individual predictions through training of a local surrogate model, called LIME (Local Interpretable Model-agnostic Explanations) [125]. For each feature, the algorithm provides a value (a score) representing the feature's importance in the resulting prediction. LIME consists of the following steps:

1. Select an instance (x, y), where y is the prediction for the sample x to explain.
2. Build a new dataset in the vicinity of instance x, perturbing the data.
3. Each point from the new dataset receives a weight according to its proximity to the instance, x.

4. Using the new weighted dataset, an intrinsic interpretable model is trained.
5. The interpretable model approximates the behavior of the non-interpretable model in the proximity of the instance x, but not global. This new interpretable model provides an explanation for the complex model's output.

We proposed a LIME-based analysis method consisting in (the method is applied in the AER model from chapter 5) [25]:

1. "LIME is run more times (in our experiments we have run it 20 times).
2. For each run, the influence of the features on the prediction is obtained (some features support the presence of a specific emotion, other features support the absence of that emotion).
3. For each feature, the number of occurrences of the feature in the subset of features, which support the presence of a specific emotion, is computed, as well as the number of occurrences of the feature in the subset of features, which support the absence of the that emotion.
4. For each feature, the absolute difference of the two frequencies from above is calculated.
5. Afterwards, the features are sorted in a descending order, based on their absolute differences; the feature with the highest absolute difference is the most contributing to predict the presence or absence of an emotion. If the number of appearances of the feature in the subset of features which supports the presence of the emotion is higher than the number of appearances of the same feature in the subset of features which supports the absence of the emotion, then we consider that the feature supports the presence of the emotion, otherwise vice versa."

Results:

We tested our analysis method based on LIME for 14 samples (7 emotions x 2 sample for each emotion). We ran LIME 20 times for a prediction and for each feature we counted the occurrences in zone (the zone contributing to the prediction of class 1 and the zone contribution to the prediction of class 0). The absolute difference of the two results shows the way in which the feature determines the predicted class. Based on the interpretations presented above, we have several situations:

- In 20 runs of the LIME algorithm all features contribute to the presence of the emotion (i.e. boredom, confusion, frustration, excitement).
- In 20 runs of the LIME algorithm some features determine more the presence of an emotion, some determine more the absence of an emotion, while some determine

equally the presence and absence of an emotion. By adding the absolute differences for the features from both categories (the presence or absence of an emotion) we obtain the results below.

Below we present the results for 2 samples, 1 for boredom absence and 1 for boredom presence. In the first row are the IDs of the features and in the second row are the associated absolute differences. With the orange colour are marked the features which contribute several times to the presence of an emotion and with blue the features which contribute several times to the absence of an emotion. The features with no colours have the absolute difference 0.

Boredom

y_predict = 0 / boredom absence

19	6	23	7	4	11	12	2	5	13	20	24
14	12	12	10	8	8	8	6	6	6	6	6

1	9	14	18	21	0	8	15	16	17	22	3	10
4	4	4	4	4	2	2	2	2	2	2	0	0

In 20 runs:

- feature 0 appears 11 times influencing the absence of boredom and 9 times influencing its presence
- feature 1 appears 12 times influencing the absence of boredom and 8 times influencing its presence
- feature 2 appears 13 times influencing the absence of boredom and 7 times influencing its presence
- feature 3 appears 10 times influencing the absence of boredom and 10 times influencing its presence
- feature 4 appears 6 times influencing the absence of boredom and 14 times influencing its presence
- feature 5 appears 13 times influencing the absence of boredom and 7 times influencing its presence
- feature 6 appears 16 times influencing the absence of boredom and 4 times influencing its presence

- feature 7 appears 5 times influencing the absence of boredom and 15 times influencing its presence
- feature 8 appears 9 times influencing the absence of boredom and 11 times influencing its presence
- feature 9 appears 12 times influencing the absence of boredom and 8 times influencing its presence
- feature 10 appears 10 times influencing the absence of boredom and 10 times influencing its presence
- feature 11 appears 14 times influencing the absence of boredom and 6 times influencing its presence
- feature 12 appears 14 times influencing the absence of boredom and 6 times influencing its presence
- feature 13 appears 13 times influencing the absence of boredom and 7 times influencing its presence
- feature 14 appears 8 times influencing the absence of boredom and 12 times influencing its presence
- feature 15 appears 11 times influencing the absence of boredom and 9 times influencing its presence
- feature 16 appears 11 times influencing the absence of boredom and 9 times influencing its presence
- feature 17 appears 9 times influencing the absence of boredom and 11 times influencing its presence
- feature 18 appears 8 times influencing the absence of boredom and 12 times influencing its presence
- feature 19 appears 3 times influencing the absence of boredom and 17 times influencing its presence
- feature 20 appears 13 times influencing the absence of boredom and 7 times influencing its presence
- feature 21 appears 12 times influencing the absence of boredom and 8 times influencing its presence
- feature 22 appears 11 times influencing the absence of boredom and 9 times influencing its presence
- feature 23 appears 4 times influencing the absence of boredom and 16 times influencing its presence

- feature 24 appears 7 times influencing the absence of boredom and 13 times influencing its presence

Summing the absolute differences, we can conclude that the features influence more the absence of boredom (72) than its presence (62).

y_predict =1 / boredom presence

3	4	23	5	7	13	15	17	19	0	14	18	20
18	18	18	16	16	16	16	16	16	13	12	12	12

22	9	16	1	6	8	10	11	21	12	2	24
12	10	10	9	8	8	8	8	8	6	4	4

In 20 runs:

- feature 0 appears 4 times influencing the absence of boredom and 17 times influencing its presence
- feature 1 appears 5 times influencing the absence of boredom and 14 times influencing its presence
- feature 2 appears 8 times influencing the absence of boredom and 12 times influencing its presence
- feature 3 appears 1 time influencing the absence of boredom and 19 times influencing its presence
- feature 4 appears 1 time influencing the absence of boredom and 19 times influencing its presence
- feature 5 appears 2 times influencing the absence of boredom and 18 times influencing its presence
- feature 6 appears 6 time influencing the absence of boredom and 14 time influencing its presence
- feature 7 appears 2 times influencing the absence of boredom and 18 times influencing its presence
- feature 8 appears 6 times influencing the absence of boredom and 14 times influencing its presence
- feature 9 appears 5 times influencing the absence of boredom and 15 times influencing its presence
- feature 10 appears 6 times influencing the absence of boredom and 14 times influencing its presence

- feature 11 appears 6 times influencing the absence of boredom and 14 times influencing its presence
- feature 12 appears 7 times influencing the absence of boredom and 13 times influencing its presence
- feature 13 appears 2 times influencing the absence of boredom and 18 times influencing its presence
- feature 14 appears 4 times influencing the absence of boredom and 16 times influencing its presence
- feature 15 appears 2 times influencing the absence of boredom and 18 times influencing its presence
- feature 16 appears 5 times influencing the absence of boredom and 15 times influencing its presence
- feature 17 appears 2 times influencing the absence of boredom and 18 times influencing its presence
- feature 18 appears 4 times influencing the absence of boredom and 16 times influencing its presence
- feature 19 appears 2 times influencing the absence of boredom and 18 time influencing its presence
- feature 20 appears 4 times influencing the absence of boredom and 16 times influencing its presence
- feature 21 appears 6 times influencing the absence of boredom and 14 times influencing its presence
- feature 22 appears 4 times influencing the absence of boredom and 16 times influencing its presence
- feature 23 appears 1 time influencing the absence of boredom and 19 times influencing its presence
- feature 24 appears 8 times influencing the absence of boredom and 12 times influencing its presence

Summing the absolute differences, we can conclude that the features influence more the presence of boredom (294) than its absence (0).

The resume of explanations for 12 samples, for which the absence or the presence of confusion, respectively frustration, curiosity, excitement, concentration and anxiety are predicted, are shown below.

Confusion

y_predict = 0/confusion absence

12	3	10	16	17	19	24	5	6	8	13	18
10	8	8	8	8	8	8	6	6	6	6	6

2	4	14	15	22	0	1	7	9	11	20	21	23
4	4	4	4	4	2	2	2	2	2	2	2	2

Summing the absolute differences, we can conclude that the features influence more the absence of confusion (94) than its presence (30).

y_predict = 1/confusion presence

1	8	9	12	13	15	18	19	20	22	23	2
20	20	20	20	20	20	20	20	20	20	20	18

3	4	6	10	11	14	16	21	24	5	17	0	7
18	18	18	18	18	18	18	18	18	16	16	14	14

Summing the absolute differences, we can conclude that the features influence more the presence of confusion (460) than its presence (0).

Frustration

y_predict = 0/frustration absence

8	6	21	0	18	7	12	1	2	5	9	11	13
16	14	12	10	10	6	6	4	4	4	4	4	4

17	4	14	16	20	23	19	22	3	10	15	24
4	2	2	2	2	2	1	1	0	0	0	0

Summing the absolute differences, we can conclude that the features influence more the absence of frustration (59) than its presence (56).

y_predict = 1/frustration presence

4	22	8	9	12	15	18	23	2	10	20	6
20	20	18	18	18	18	18	18	17	16	16	14

7	16	19	0	14	17	24	1	3	11	21	5	13

14	14	14	13	12	12	12	10	10	10	10	8	8

Summing the absolute differences, we can conclude that the features influence more the presence of frustration (358) than its presence (0).

Curiosity

y_predict = 0/curiosity absence

11	21	3	6	1	9	12	18	2	17	23	7	10
17	16	13	11	9	9	9	9	7	7	7	5	5

13	0	5	14	4	8	15	16	19	20	22	24
5	3	3	3	1	1	1	1	1	1	1	1

Summing the absolute differences, we can conclude that the features influence more the absence of curiosity (115) than its presence (31).

y_predict =1/curiosity presence

6	13	4	19	24	8	10	15	16	20	3	0
16	16	14	14	14	12	12	12	12	12	11	10

5	22	9	17	21	23	7	14	2	11	12	1	18
10	10	9	8	8	8	6	6	2	2	2	0	0

Summing the absolute differences, we can conclude that the features influence more the presence of curiosity (224) than its presence (2).

Excitement

y_predict = 0/excitement absence

2	7	21	22	14	1	4	5	6	11	17	20	23
14	14	12	12	10	8	6	6	6	6	6	6	6

3	8	19	9	12	13	15	16	0	10	18	24
4	4	4	2	2	2	2	2	0	0	0	0

Summing the absolute differences, we can conclude that the features influence more the absence of excitement (80) than it's presence (54).

y_predict =1/excitement presence

4	10	16	2	6	8	17	18	20	14	22	24	13
20	20	20	18	18	18	18	18	18	17	16	16	14

19	21	23	9	1	3	15	0	11	5	12	7
14	14	14	13	12	12	12	10	10	8	8	4

Summing the absolute differences, we can conclude that the features influence more the presence of excitement (362) than it's presence (0).

Concentration

y_predict = 0/concentration absence

11	16	17	1	8	6	19	20	5	18	9	0	4	7
20	20	18	16	16	14	10	10	8	8	6	4	4	4

10	12	14	2	3	13	23	24	15	21	22
4	4	4	2	2	2	2	2	0	0	0

Summing the absolute differences, we can conclude that the features influence more the absence of concentration (102) than its presence (78).

y_predict =1/concentration presence

11	6	8	15	16	20	7	0	17	2	4	5	9	12
20	18	18	16	16	16	14	12	12	10	10	10	10	10

22	23	3	10	19	21	14	24	1	13	18
10	10	8	8	8	8	6	6	4	4	0

Summing the absolute differences, we can conclude that the features influence more the presence of concentration (216) than its presence (48).

Anxiety

y_predict = 0/anxiety absence

8	11	12	16	18	2	1	3	4	22	6	14	17
17	17	17	17	15	13	11	11	9	9	7	7	7

20	24	7	9	13	15	0	19	21	23	5	10
7	6	5	5	5	5	3	3	3	3	1	1

Summing the absolute differences, we can conclude that the features influence more the absence of anxiety (149) than its presence (55).

y_predict =1/anxiety presence

17	18	20	0	6	10	16	11	21	22	2	5	14
18	16	16	14	14	14	14	12	12	12	10	10	10

23	24	7	13	19	1	3	12	8	15	4	9
10	10	8	8	6	4	4	4	2	2	0	0

Summing the absolute differences, we can conclude that the features influence more the presence of anxiety (210) than its presence (20).

A cluster-based aggregation method for aligning explanations

To solve the disagreement problems between various explainers, we proposed a method inspired by Case-Based Reasoning to aggregate different explanations [30]. In our research we considered the Pirie et al. (2023) approach, which applied the local alignment and proposed Case Alignment Confidence metric between explainers and developed a framework for explainers' aggregation, named AGREE (AGgregation for Robust Explanations) [35].

The feature attribution vectors containing explanations (feature importance scores and signs) represent the case bases for our algorithm. We generated feature attribution vectors using LIME, SHAP (Kernel SHAP and Tree SHAP) and Anchors algorithms, for the following datasets: Pima Indian Diabetes Dataset [36], Indian Liver Patient Dataset [37], Hepatitis Dataset [38], Fetal Dataset [39], Abalone Dataset [40], Water Quality Dataset [41].

Our method consisted in two stages. In the first stage, there were generated the explanations and in the second stage we applied a cluster-based aggregation method inspired from Case-Based Reasoning. In our approach, feature attribution vectors represented the case bases.

In stage 1 we used 100 times a Leave-One-Out Cross-Validation procedure and computed the atribution vectors for each explainer (LIME, Anchors, Kernel SHAP, and Tree SHAP). In the case of LIME and SHAP, the atribution vectors contains the atribution scores and sign, whereas in the case of Anchors the vectors contain only the attribution scores (Figure 6.2. and Figure 6.3.). For example, for Diabetes dataset we obtained 76800 (768 instancesx100) attribution vectors for each explainers.

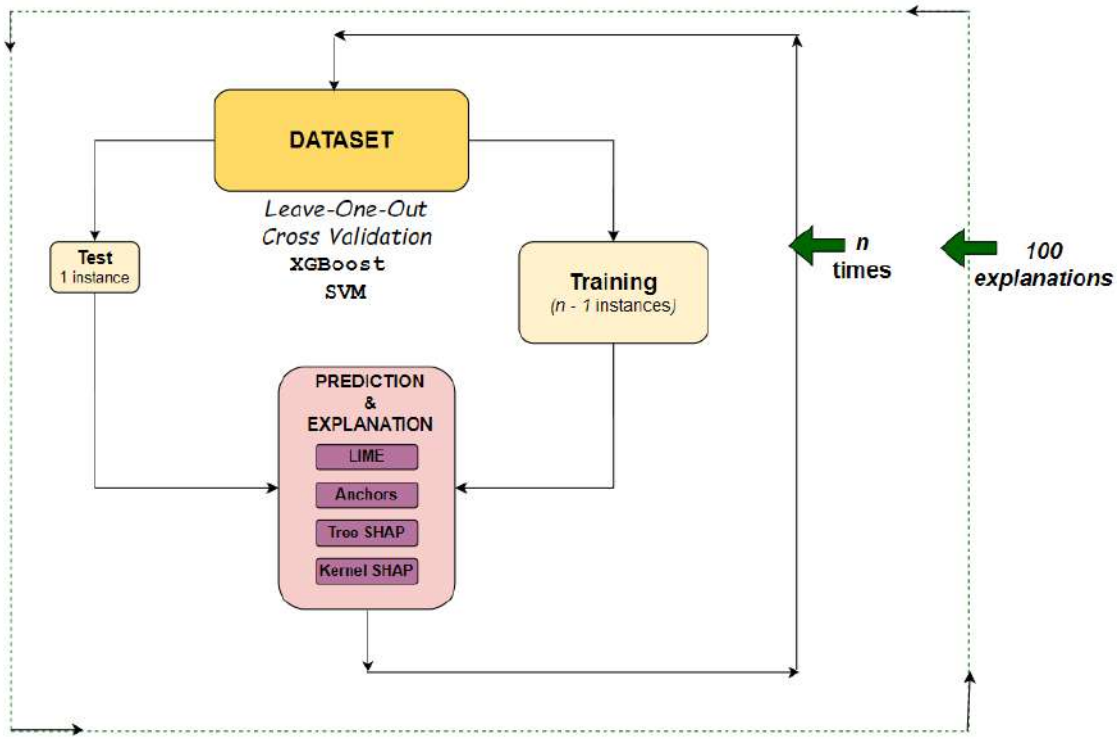

Figure 6.2. Leave-One-Out Cross Validation procedure for generating explanations [30]

```
for each explanation method E in [LIME, Anchors, TreeSHAP, KernelSHAP]:
    for each dataset D in [Diabetes, Liver, Hepatitis, Abalone, Water, Fetal]:
        for each instance i from D:
            for a number of 100 iterations:
                apply Leave-One-Out Cross-Validation

                if E == LIME or E == TreeSHAP or E == Anchors:
                    model = XGBoost classifier
                    generate prediction
                else if E == KernelSHAP:
                    model = Support Vector Machine
                    generate prediction
                if E == LIME OR E == TreeSHAP or E == KernelSHAP:
                    generate explanation
                    feature attribution vector = attribution scores + signs
                else if E == Anchors:
                    generate explanation
                    feature attribution vector = attribution scores

                calculate vector of feature attribution ranks R
                extract vector of feature attribution signs S
                extract vector of feature attribution ranks and signs SR
```

Figure 6.3. Pseudocode for the algorithm that generates explanations and feature attribution vectors [30]

In stage 2, there were computed alignment scores between problem and solution clusters of explainers and explanations using a complexity measure, namely GAME (Global Alignment MEasure) proposed by Chakraborti et al. (2007) [126].

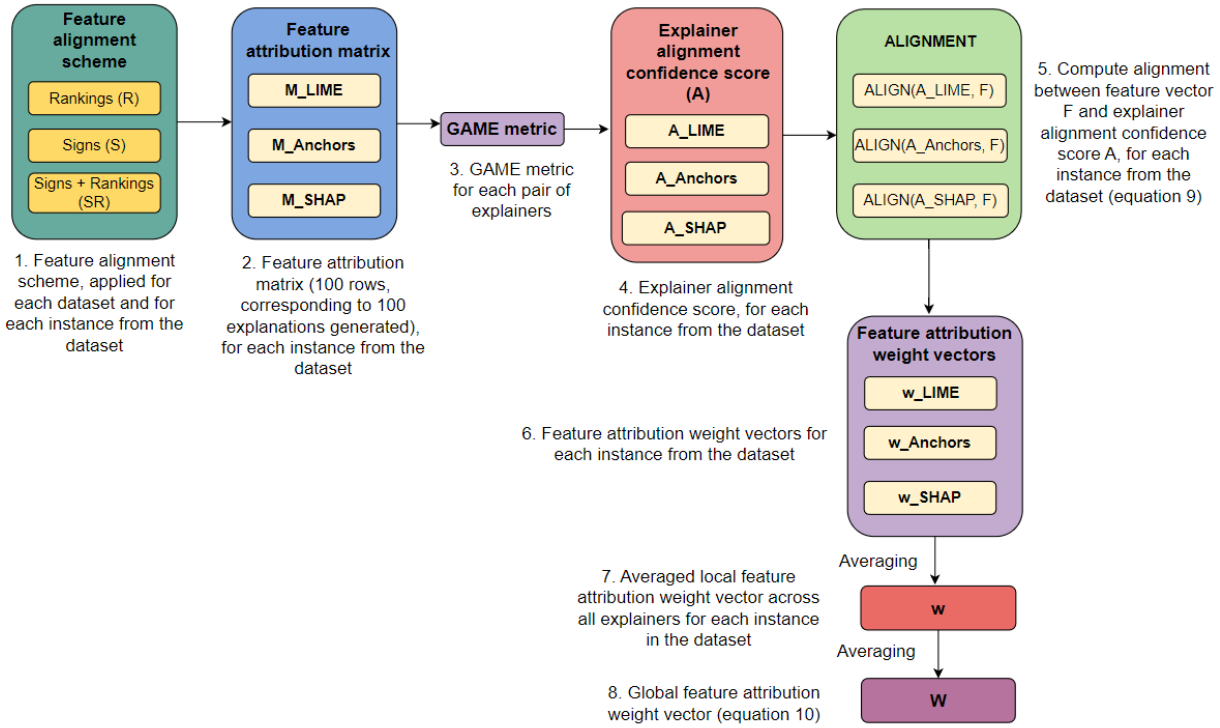

Figure 6.4. Aggregation strategy workflow[30]

For feature attribution ranks (R feature scheme), for each instance from the dataset, we generated the matrices M_Anchors, M_LIME, and M-SHAP. Each matrix had 100 rows, corresponding to the number of iterations from scheme 6.2. Tree SHAP and Kernel SHAP was stable and consistent, compared to LIME and Anchors, which generated unstable explanations. So, M_SHAP matrix contained 50 explanations from Tree SHAP and 50 explanations from Kernel SHAP. The matrices M_Anchors, M_LIME and M_SHAP were aligned using the cluster-based Global Alignment approach. We computed the GAME metric for each pair Anchors-Anchors, Anchors-LIME, Anchors-SHAP, LIME-Anchors, LIME-LIME, LIME-SHAP, SHAP-Anchors, SHAP-LIME, SHAP-SHAP. Obviously, the GAME metric for Anchors-Anchors, LIME-LIME, SHAP-SHAP was 1.

In the cases of feature attribution signs and feature attribution signs + ranks scheme (S and SR), we computed only the matrices M_LIME and M_SHAP and GAME metric for the pairs LIME-LIME, LIME-SHAP, SHAP-LIME and SHAP-SHAP. Using the GAME scores between explainers, we calculated an explainer confidence vector containing a confidence score for each explainer.

Finally, for each scheme R, S and SR, for each instance, we computed the explainer confidence. Using the explainer confidence and feature attribution rank/feature attribution signs/both feature attribution ranks and feature attribution signs, we obtained the consensus feature attribution weight vector for each explainer. Averaging all the local weight

vectors, we calculated the global feature attribution weight vector. The pseudocode of the algorithm is presented in Figure 6.5.

```
for each feature scheme F in [R, S, SR]:
    for each dataset D in [Diabetes, Liver, Hepatitis, Abalone, Water, Fetal]:
        for each instance i from D:
            if F == R:
                generate matrices M_LIME, M_Anchors, M_SHAP
                for each matrix Mi in [M_LIME, M_Anchors, M_SHAP]:
                    for each matrix Mj in [M_LIME, M_Anchors, M_SHAP]:
                        calculate GAME(Mi, Mj)
            if F == S or F == SR:
                generate matrices M_LIME, M_SHAP
                for each matrix Mi in [M_LIME, M_SHAP]:
                    for each matrix Mj in [M_LIME, M_SHAP]:
                        calculate GAME(Mi, Mj)

            calculate explainer confidence A
            for each explainer E:
                calculate local feature attribution weight vectors wbaraE
            calculate averaged local feature attribution weight vector wbara across all explainers E in []
        calculate averaged global feature attribution weight vector W across all instances
```

Figure 6.5. The pseudocode for the algorithm that aligns the explainers and explanations [30]

We evaluated the proposed strategy by providing the aggregated explanation weight vectors to the feature space of a weighted k-NN classifier (for each alignment scheme R, S, SR) and comparing the prediction performance against a non-weights k-NN algorithm, having as task a binary classification (Figure 6.6). Moreover, we performed comparations between the weighted k-NN algorithm having the aggregated feature overlap explanation weights with the weighted k-NN algorithm having the weights produced by LIME, SHAP or Anchors.

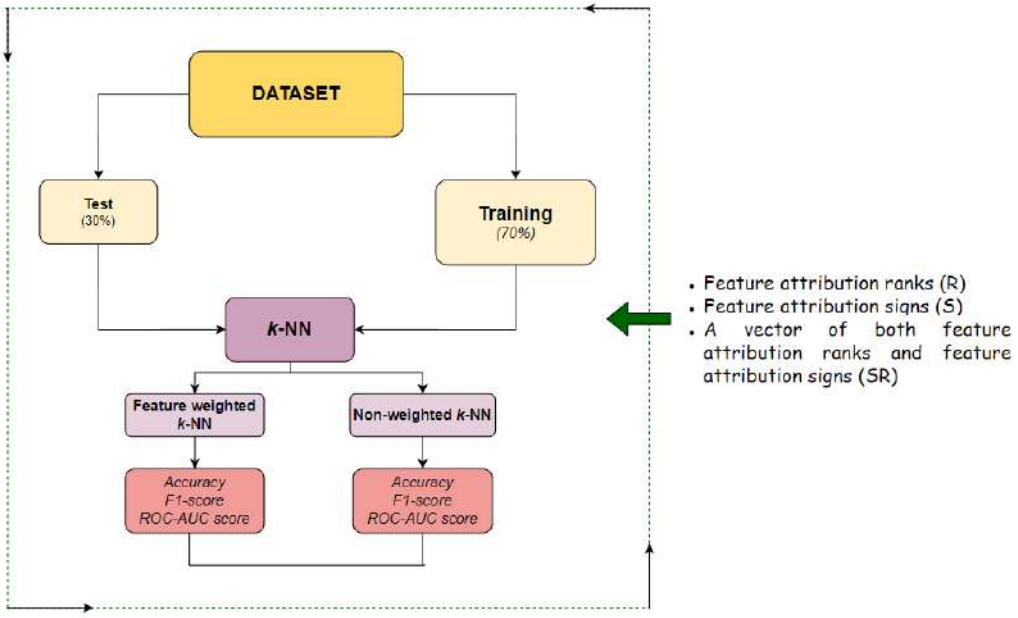

Figure 6.6. Evaluation of the feature weighted and non-weighted k-NN classifiers [30]

Results:

We applied our method on 6 popular databases listed above and aligned multiple explanations by comparing problem and solution space clusters. We performed more

experiments using 9 feature overlap explanation scheme for explainers and explanations alignment: R_AVG – the mean of rankings, R_FI – the mean of feature importances as obtained from a XGBoost classifier, R_AVG_FI – the mean of rankings multiplied by the mean of feature importances, R_A – the mean of rank alignments obtained from the Global Alignment Measurement, R_A_FI – the mean of rank alignments obtained from the Global Alignment Measurement multiplied by the mean of feature importances as obtained from a XGBoost classifier, S_A – the mean of sign alignments obtained from the Global Alignment Measurement, S_A_FI - the mean of sign alignments obtained from the Global Alignment Measurement, SR_A – the mean of the vector containing rank and sign alignments obtained from the Global Alignment Measurement, SR_A_FI - the mean of the vector containing rank and sign alignments obtained from the Global Alignment Measurement, multiplied by the mean of feature importances as obtained from a XGBoost classifier.

For each alignment scheme R, S, SR, we compared the performance of feature weighted k-NN against the performance of non-weighted k-NN.

Having the explanation weights $\overline{w}$ provided by a feature overlap method, the weighted Euclidian distance between two instances, x and y, in the k-NN algorithm was calculated as follows:

$$dist(x,y) = \sqrt{\sum_{i=1}^{m} \overline{w}(x_i - y_i)^2}$$

, where m represents the number of features for each instance.

The weighted and the non-weighted k-NN classifiers were run 50 times and the results were averaged. We evaluated the models' performances using accuracy, F1-score and ROC-AUC as metrics. The pseudocode for the evaluation of the weighted k-NN and non-weighted k-NN is shown in Figure 6.7.

```
split the dataset into 30% test and 70% training
if F == R:
    for each feature overlap explanation weights method FW in [R_AVG, R_FI, R_AVG_FI, R_A, R_A_FI]:
        apply Feature weighted k-NN using FW
        calculate accuracy, F1-score, ROC-AUC score
        apply Non-weighted k-NN
        calculate accuracy, F1-score, ROC-AUC score
        compare results
if F == S:
    for each feature overlap explanation weights method FW in [S_A, S_A_FI]:
        apply Feature weighted k-NN using FW
        calculate accuracy, F1-score, ROC-AUC score
        apply Non-weighted k-NN
        calculate accuracy, F1-score, ROC-AUC score
        compare results
if F == SR:
    for each feature overlap explanation weights method FW in [SR_A, SR_A_FI]:
        apply Feature weighted k-NN using FW
        calculate accuracy, F1-score, ROC-AUC score
        apply Non-weighted k-NN
        calculate accuracy, F1-score, ROC-AUC score
        compare results
```

Figure 6.7. The pseudocode for the evaluation of the weighted k-NN and non-weighted k-NN [30]

The tables with the results are presented in the appendix of the paper [30]. Here we present a short resume of them.

The performance scores for weighted k-NN algorithm were greater than the performance scores obtained with non-weighted k-NN classifier for the Diabetes, Liver, Hepatitis, Water, and Fetal datasets. The exception made the Abalone dataset, for which we obtained lower classification scores. In Figure 6.8, we present the increase of the averaged weighted k-NN performance metric scores (accuracy, F1, ROC-AUC) from the non-weighted k-NN performance scores. Also, in tables 6.1-6.3. we highlight the values for each metrics, dataset and overlap explanation.

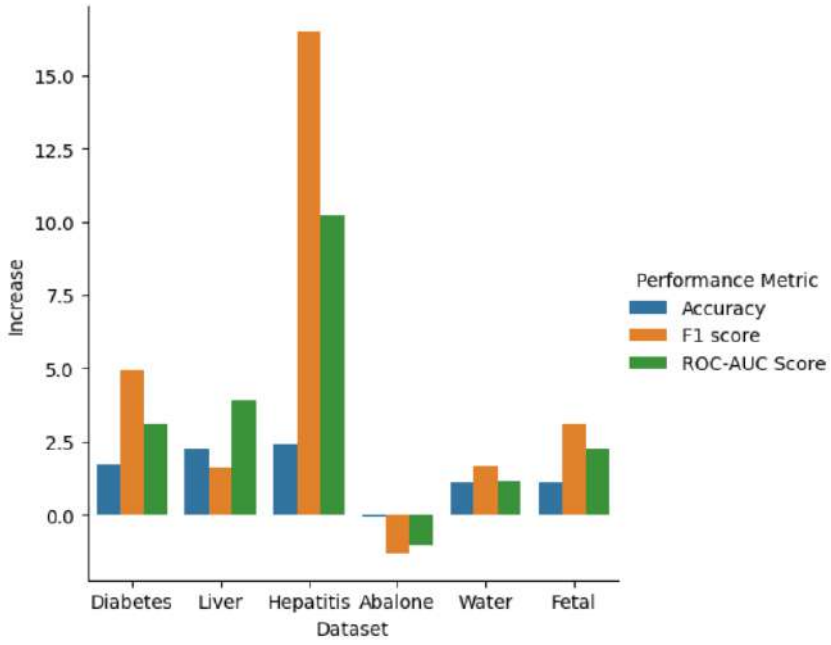

Figure 6.8. The increase of the weighted k-NN performance from non-weighted k-NN for each dataset [30]

Comparison between the averaged metrics of the weighted k-NN predictions using each of the 9 feature overlap explanation weights to the predictions of the non-weighted k-NN classifier (Table 6.1, 6.2).

Table 6.1 Accuracy scores of the weighted k-NN algorithm comparing with the non-weighted k-NN classifier [30]

	Accuracy (%)								
Dataset	R_AVG	R_FI	R_AVG_FI	R_A	R_A_FI	S_A	S_A_FI	SR_A	SR_A_FI
Diabetes	74.97	74.13	74.7	74.17	**75.13**	-	73.5	73.77	74
Liver	68.59	-	**69.04**	67.90	68.66	67.26	67.3	67.12	67.57
Hepatitis	-	95.49	**95.54**	-	95.50	93.3	94.43	-	94.84
Abalone	-	**54.47**	52.67	-	-	-	-	-	-
Water	61.40	62.33	62.14	61.86	**62.37**	61.47	62.23	-	61.98
Fetal	92.51	92.82	93.47	92.53	**93.52**	92.2	92.94	-	92.96

	Diabetes	Liver	Hepatitis	Abalone	Water	Fetal
Average accuracy for weighted k-NN	**74.3**	**67.93**	**94.85**	**53.55**	**61.97**	**92.87**
Non-weighted k-NN	72.58	65.65	92.43	53.62	60.86	91.74

Table 6.2 F1 scores of the weighted k-NN algorithm comparing with the non-weighted k-NN classifier [30]

	F1 score (%)								
Dataset	R_AVG	R_FI	R_AVG_FI	R_A	R_A_FI	S_A	S_A_FI	SR_A	SR_A_FI
Diabetes	61.95	59.86	62.18	61.33	**62.79**	-	57.95		58.64
Liver	78.46	-	**78.57**	77.84	78.42	77.98	77.87	77.59	77.93
Hepatitis	-	77.94	**79.07**	62.17	78.6	63.65	74.52	61.38	74.7
Abalone	-	**46.44**	-	-	-	-	-	-	-
Water	43.72	44.56	44.4	44.2	**45.1**	-	44.5	-	44.8
Fetal	82.15	82.91	84.79	82	**84.82**	81.11	83.41	-	83.17

	Diabetes	Liver	Hepatitis	Abalone	Water	Fetal
Average F1 score for weighted k-NN	**60.67**	**78.08**	**71.51**	**46.44**	**72.35**	**83.04**

Non-weighted k-NN	55.74	76.45	55.01	47.78	42.81	79.95

Table 6.3 ROC-AUC scores of the weighted k-NN algorithm comparing with the non-weighted k-NN classifier [30]

	ROC-AUC score (%)								
Dataset	R_AVG	R_FI	R_AVG_FI	R_A	R_A_FI	S_A	S_A_FI	SR_A	SR_A_FI
Diabetes	71.28	69.92	71.35	70.69	**71.93**	-	68.66	-	69.15
Liver	59.88	-	**61.4**	59.63	60.39	-	-	-	-
Hepatitis	73.45	83.82	**85.02**	-	84.54	73.89	82.83	72.91	81.81
Abalone	-	**52.07**	-	-	-	-	-	-	-
Water	57.28	58.20	57.92	57.70	**58.28**	57.293	58.04	-	58
Fetal	87.17	88.09	89.01	87.05	**89.35**	-	88.19	-	88.06

	Diabetes	Liver	Hepatitis	Abalone	Water	Fetal
Average ROC-ACU of weighted k-NN	**70.42**	**60.33**	**79.78**	**52.07**	**57.84**	**88.13**
Non-weighted k-NN	67.34	56.41	69.59	53.13	56.68	85.88

From the data presented above, we can resume:

- R_A_FI contributed to the highest accuracies, F1-scores and ROC-AUC score in the cases of Diabetes, Water and Fetal datasets.
- R_AVG contributed to the highest accuracies, F1-scores and ROC-AUC score in the cases of Liver and Hepatitis datasets.
- R_FI contributed to the highest accuracies, F1-scores and ROC-AUC score in the cases of Abalone dataset.

The results show that that R_AVG_FI and R_A_FI are the feature overlap explanation weight methods with the highest impact on explainers and explanations consensus [30].

(B-ii) The evolution and development plans for career development

This chapter aims to present future directions of research, didactic evolution, and academic community development plans.

The research evolution plan

- Main research directions

The author of this habilitation thesis will continue to expand the machine learning research group by involving more students and furthering the collaborations in academia and industry by establishing partnerships on research topics of common interest.
I am currently concerned on three main research directions **1. low-power machine learning models, 2. explainability in machine learning** and **3. ethics in machine learning**.
Below, I make a brief presentation of these subjects, which were also addressed in this thesis.

Low-power machine learning algorithms and affective computing

In the recent research report [130], it shows that in 2025 the electricity demand for AI data centers will likely reach 21 GW, twice as much compared to 2023. In 2027, it is estimated that a data center hosting AI training will require the energy generated by a nuclear power plant.
The expansion of Generative AI requires more power both for specific algorithms training and for usage of it. In [131], the authors argue the importance of a responsible and sustainable development of the Gen-AI sector. They suggest the usage of a framework to perform a benefit-cost evaluation to "encourage (or require) Gen-AI to develop in ways that support social and environmental sustainability goals alongside economic opportunity".
To decrease the consume of the energy of AI, the neuromorphic engineering proposes to use "the computational principles of the brain" [132]. Some promising low-power algorithms are those based on spiking neurons.
The networks of spiking neurons (SNN) have the same architecture as artificial neural networks differing through the computational units, called spiking neurons [132]. Information is coded in spikes and is transmitted from one neuron to another through

synapses. The inputs for a neuron are spikes, very short bursts of electrical activity. Thus, the potential membrane of the neuron increases by accumulating the spikes. When the potential of the neuron's membrane reaches a threshold, the neuron generates a spike to the neurons connected to it. These neurons accumulate spikes until the threshold value is reached. Basically, they wait for enough inputs to accumulate to trigger spikes. After the neuron fires a spike, the membrane potential is reset.

There are more models of spiking neurons depending on the spike generation mechanism [132], [133]. The most popular models are the leaky integrate and fire (LIF) model based on Lapicque's model [134] and the spike-response model (SRM), a generalization of the LIF [135], [136].

In LIF model, the membrane of a neuron is modelled through a differential equation:

$$\tau \frac{du(t)}{dt} = u_{rest} - u(t) + RI(t), u < u_{th}$$

where u is the membrane potential, u_{rest} is the resting potential, τ is the membrane time constant, R the membrane resistance, I represent the input, the weighted sum of spikes, u_{th} is the firing threshold.

We used SNNs with LIF model for the seven emotions recognition from chapter 5 (boredom, confusion, frustration, curiosity, excitement, concentration, anxiety) and obtained good results in terms of accuracy and running time (Table 1).

Table 1. Running time for the recognition of seven emotions with SNN

Emotion	SNN - Performance (%)		
	Training accuracy (%)	Test Accuracy (%)	Running time (seconds)
Boredom	97.656%	97.861%	153.5725
Confusion	100%	96.378%	142.9799
Frustration	95.312%	96.953%	**192.3776**
Curiosity	98.4375%	96.969%	128.8649
Excitement	98.437%	98.672%	110.2392
Concentration	93.75%	91.798%	122.2810
Anxiety	94.5312%	97.417%	121.9561

Comparing the running time for training we obtained a sharp decrease of it, the maximum running time for SNN training is **192.37** seconds, while the minimum time for the Model 1D-CNN with 5 EEG is **5790.36** seconds.

Regarding the affect aware applications, we will focus on adapting SNNs for emotion recognition both from biophysical signals and facial expression. For emotion recognition from facial expression, we will use AffectNet dataset, for which we have received rights of use for research purposes [137].
A person feels a mixture of emotions at a given moment. We intend to buid AERs based on multilabel classification through which to relieve the range of emotions felt by a person at a given moment.
Also, we intend to develop explanations algorithms for SNN-based predictions. This topic is relatively new [138], [139] and we intend to provide reliable explanations specific for SNNs. Also, the author intends to include PhD students in this work.

Explainability in machine learning

The topic of explainability in machine learning is a challenge. We must face the disagreement problem and the requirements of high computational resources.
Some aspects of the disagreement problem in XAI are presented in Chapter 6 of this thesis. We consider that it is a critical issue, which requires a rapid solution so that people can trust AI. I and my colleagues are currently working on a new reliable explanation LIME-based algorithm and on an algorithm to obtain meta-ranks for features. The results obtained so far are promising, which led the team to continue the research with experimental studies.
The perturbation-based methods used to generate explanations for an opaque model are computationally expensive and can generate inconsistent results. Also, global methods suffer from the same inconveniences. Moreover, there are applications based on ML models that need explanations in real-time. The scope of the research focusses currently to propose improvements of explanations methods to assure both the real-time and trusted explainability. For example, in [140], a content-addressable memory (CAM)-based architecture is defined to accelerate the tree-based ML models.
Our goal is also the leveraging between high performant ML models and low-resource consistent explanations methods. This is a hot topic, and I plan to involve PhD students in the subject's research.

Ethics in machine learning

Ethics in machine learning is a sensitive and controversial topic. The AI and ML researchers play a significant role in ensuring ethical compliance. "Being close to the technology, AI/ML researchers are well placed to highlight new risks, develop technical solutions, and choose to work for organizations that align with their values" is shown in [141]. More and more, the definition and regulation of ethical principles are being considered, as happened in Europe with the AI Act [4]. We must address and to mitigate the potential bias identified in [142] which could occurs in AI/ML models: data bias, development bias and interactions bias. AI ethics is defined in [29] as "a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies". AI ethics involves not only fairness, accountability, and transparency of the algorithms, but also sustainability of the models. I want to continue to develop ethical machine learning models, a subject of great interest nowadays.
Currently, I am member of a new project entitled Strategies for Creating Equitable Workplaces in Society 5.0 and I am working with a masters' student on a ML-based solution for information extraction and processing from pdf files stored in open databases. This is shaping a new direction of research, on the development of an ML-based assistant for GDPR compliance in the employment relationships.

On these topics and other related to, the author will continue the collaboration with the colleagues from the National University of Science and Technology POLITEHNICA of Bucharest.

- Disseminate the research results

The author will continue to publish the research results in highly ranked journals and conference proceedings. I will collaborate with colleagues both from my university and other universities. The PhD students will be engaged in all aspects of the research activities.

- Apply for research projects

The author will identify the calls for the national and EU research projects on subjects of interest and apply to obtain funds to support the research activities of our team, future PhD

students, our department and university. To have a successful application, there will be extended the collaborations with CS departments from other universities and with ICT companies.
Our university organizes an annual competition for research projects, in which I have participated and won in 2013 a project entitled Affective Learning: Benefits and Ethical Risks in Higher Education (ALBER). I will continue to apply to these calls, which provide funds for conference attendance, equipment purchases and salary payments for project staff.

- Collaboration with universities and companies

I have collaborated with colleagues from many universities doing research, writing papers or projects proposals. Being responsible on behalf of the Petroleum-Gas University of Ploiesti for the partnership contracts between the university and 7 companies, the author will maintain these commitments and extend the research collaborations.
I teach "The research activities in Computer Science" course and every year there are invited PhD supervisors from our universities and abroad to give lectures. Also, I have invited specialists in the field to present the latest trends in CS research to colleagues from our university. I plan to continue these collaborations across the academic and industry.

- Organize the scientifical workshops and seminars

I am an organizing member of The Computer Science and Education Symposium and Scientific Workshop Trends in the Computer Science Research for students in the master, to which I will continue to contribute.
Having the experience of organizing the MAI-XAI 24 workshop, I will continue to propose workshops at conferences and to get involved in the organization of these workshops.

- Serving in PC for conferences

For more than 5 years, I have been serving on the program committee for the ACM ITICSE conference. I intend to continue the work at this conference, from which I have learned to manage the demanding process of reviewing papers for a highly ranked conference. Also, I am member in PC for other conferences and I will continue to contribute to their organization. I reviewed many papers from prestigious journals such as IEEE Access,

Information Fusion, IEEE Journal of Biomedical and Health Informatics, Scientific Reports, Cognitive Systems Research. I will continue the reviewer activity for scientific journals and conferences.

- Develop the Computer Science&Education Laboratory and research laboratories from my department, Petroleum-Gas University of Ploiești

In 2020, I won and managed the project entitled POLE4R&D – Support for Research in STEM Research, Support axis (CNFIS FDI) through which I set up the Math Research Pole, coordinated by a colleague of mine. In 2019, another project was won, the project entitled UPG-HUB4.0 - Multidisciplinary Research, Development and Innovation Hub, Research Support axis (CNFIS FDI). Through it, the UPG HUB and the Computer Science&Education Laboratory have been set up. I coordinate the Computer Science&Education Laboratory, which I intend to develop, involving more colleagues in it and proposing new research topics. Also, using funds from different projects, which I have coordinated, it was established and developed the department's library, and I hope to continue this task.

The didactic evolution and academic community development plan

The didactic activity represents also an important part of the author's academic career.
At present, the author has the following undergraduate and master's programs in computer science:

- Machine Learning, Data analysis, Cryptography and Information Security, Graph Algorithms, Computer Network and Research Activities in Computer Science.

I update my courses periodically with new topics and examples. I use various teaching strategies (project based, group discussion, flipped classroom, case studies, brainstorming, peer teaching, and so on) to meet the students' needs and to engage students in the learning process. The results of the various strategies that my colleagues and I apply can be seen both in the students' academic grades and in the rankings, they achieve in the national competitions and conferences they take part in, for example UNbreakable Romania 2023, 2024, The International Conference on Applied Informatics Imagination, Creativity, Design, Development, Sibiu.
I am the coordinator of the Computer Science undergraduate program. Periodically, I organize brainstorming sessions inviting students, academic staff, industry stakeholders,

and alumni to discuss the improvements that can be made to the study program. Also, I invite representatives from the IT industry to hold lectures for students to ease the transition to the job market. I coordinate the extracurricular projects entitled Computer Science, a successful career! which provides me with funds to organize students' events. I am concerned about the career orientation of students, and I try to give them the best advice to develop both professionally and personally.

I am involved in the Computer Science bachelor and master study programs promotion, and I organize regular meetings with high school students to introduce them to our study programs.

I encourage students to participate in Erasmus mobility programs and I have some recent success in this respect.

I coordinate undergraduate and master's students in the realization of their graduation thesis, as well as high school teachers for obtaining the first teaching degree.

I will continue all these activities, and I will improve them.

Since 2016, I am an evaluator of The Romanian Agency of Quality Assurance in Higher Education, Informatics field. I have a good collaboration with the members of the Informatics committee, and I will continue collaborations for the benefit of computer science education in our country.

I constantly listen to and participate in lectures given by industry specialists, considering that in the IT field there must be very close links between academia and industry.

To summarize, the following directions will be followed in the didactic evolution and academic community development:

- Constantly monitoring the study programme I coordinate. Through consultations with students, fellow teachers, researchers, and industry professionals, as well as participating in various professional organizations, I will ensure putting the student first in the educational process. Further than that, I will realize educational plans compatible with the industry requirements at a national and European level.
- Updating the courses' curriculum with the newest results of related research.
- Developing the cooperation between the department and the socio-economic and cultural sector by inviting specialists to hold lectures to students, by establishing collaborations between the department and relevant companies and research institutions etc.
- Promoting the study programme through actions both at a university level and at highschool level.

- Digitalizing the study materials.
- Introducing elective courses related to the requirements of the work field.
- Developing the department's library.
- Analysing the course content and correlating it with recommendations from prestigious international professional organizations and other curricula from relevant universities at a national and international level.
- Supporting research initiatives and including them and their results in the learning process at both an undergraduate and graduate level.
- Organizing workshops where teaching staff can present their research to students and fellow colleagues.
- Increasing the complexity of the diploma projects, encouraging research topics with an interdisciplinary approach, which have direct impact on industry, education, economy, culture, governance etc.
- Organizing scientific events dedicated to students along with extracurricular activities
- Providing career counselling to students.

(B-iii) Bibliography

[1] Picard, R., Affective Computing, M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 321, Cambridge, 1995.

[2] Picard, R.W., Papert, S., Bender, W. Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy D., and Strohecker, C.: Affective Learning – A Manifesto, BT Technology Journal, 22: 253, 2004. https://doi.org/10.1023/B:BTTJ.0000047603.37042.33.

[3] . D'Mello, S. K., & Graesser, A. C., Feeling, thinking, and computing with affect-aware learning technologies. In R. A. Calvo, S. K. D'Mello, J. Gratch, & A. Kappas (Eds.), The Oxford handbook of affective computing (pp. 419–434). Oxford University Press. 2015. https://doi.org/10.1093/oxfordhb/9780199942237.013.032

[4] Artificial Intelligence Act, Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828.

[5] Bălan, O., **Moise, G.**, Moldoveanu, A., Moldoveanu, F. and Leordeanu. M., Does automatic game difficulty level adjustment improve acrophobia therapy? differences from baseline. In Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18). Association for Computing Machinery, New York, NY, USA, Article 78, 1–2. 2018., https://doi.org/10.1145/3281505.3281583.

[6] Bălan, O., **Moise, G.**, Moldoveanu, A., Leordeanu, M., Moldoveanu. F., Challenges for ML-based Emotion Recognition Systems in Medicine. A Human-Centered Approach. CHI'19 Extended Abstracts, May 4-9, 2019, Glasgow, Scotland, UK. ACM CHI Conference on Human Factors in Computing Systems Workshop on Emerging Perspectives in Human-Centered Machine Learning.

[7] Bălan, O., Cristea, Ș., Moldoveanu, A., **Moise, G.**, Leordeanu, M., Moldoveanu, F. (2020). Towards a Human-Centered Approach for VRET Systems: Case Study for Acrophobia. In: Siarheyeva, A., Barry, C., Lang, M., Linger, H., Schneider, C. (eds) Advances in Information Systems Development. ISD 2019. Lecture Notes in Information Systems and Organisation, vol 39. Springer, Cham. https://doi.org/10.1007/978-3-030-49644-9_11.

[8] Bălan, O., Moldoveanu, A., Petrescu, L., **Moise, G.**, Cristea, Ș., Petrescu, C., Sensors system methodology for artefacts identification in Virtual Reality games, 2019 International

Symposium on Advanced Electrical and Communication Technologies (ISAECT), Rome, Italy, 2019, pp. 1-6, doi: 10.1109/ISAECT47714.2019.9069719.

[9] Balan, O., **Moise, G.**, Moldoveanu, A., Moldoveanu, F. and Leordeanu, M., Automatic Adaptation of Exposure Intensity in VR Acrophobia Therapy, Based on Deep Neural Networks. In Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden, June 8-14, 2019. ISBN 978-1-7336325-0-8 Research Papers. https://aisel.aisnet.org/ecis2019_rp/52.

[10] Bălan, O.; **Moise, G.**; Moldoveanu, A.; Leordeanu, M.; Moldoveanu, F. An Investigation of Various Machine and Deep Learning Techniques Applied in Automatic Fear Level Detection and Acrophobia Virtual Therapy. Sensors 2020, 20, 496. https://doi.org/10.3390/s20020496.

[11] Bălan, O., **Moise, G.**, Moldoveanu, A., Moldoveanu, F., Leordeanu, M., Classifying the Levels of Fear by Means of Machine Learning Techniques and VR in a Holonic-Based System for Treating Phobias. Experiments and Results. In: Chen, J.Y.C., Fragomeni, G. (eds) Virtual, Augmented and Mixed Reality. Industrial and Everyday Life Applications. HCII 2020. Lecture Notes in Computer Science, vol 12191. Springer, Cham. 2020. https://doi.org/10.1007/978-3-030-49698-2_24

[12] Petrescu, L.; Petrescu, C.; Mitruț, O.; **Moise, G.**; Moldoveanu, A.; Moldoveanu, F.; Leordeanu, M. Integrating Biosignals Measurement in Virtual Reality Environments for Anxiety Detection. Sensors 2020, 20, 7088. https://doi.org/10.3390/s20247088.

[13] Koelstra, S.; Muehl, C.; Soleymani, M.; Lee, J.-S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. DEAP: A Database for Emotion Analysis using Physiological Signals. IEEE Trans. A_ect. Comput. 2012. 3, 18–31, doi: 10.1109/T-AFFC.2011.15.

[14] Bălan, O.; **Moise, G.**; Moldoveanu, A.; Leordeanu, M.; Moldoveanu, F. Fear Level Classification Based on Emotional Dimensions and Machine Learning Techniques. Sensors 2019, 19, 1738. https://doi.org/10.3390/s19071738

[15] Petrescu, L.; Petrescu, C.; Oprea, A.; Mitruț, O.; **Moise, G.**; Moldoveanu, A.; Moldoveanu, F. Machine Learning Methods for Fear Classification Based on Physiological Features. Sensors 2021, 21, 4519. https://doi.org/10.3390/s21134519.

[16] Bălan, O.; **Moise, G.**; Petrescu, L.; Moldoveanu, A.; Leordeanu, M.; Moldoveanu, F. Emotion Classification Based on Biophysical Signals and Machine Learning Techniques. Symmetry 2020, 12, 21. https://doi.org/10.3390/sym12010021.

[17] Todorovska, E., 22 Astonishing Phobia Statistics for 2024, 22 Astonishing Phobia Statistics for 2024, https://medalerthelp.org/blog/stats-and-facts/phobia-statistics/, Accessed March 2025.

[18] North, M.M., North, S.M., and Joseph R. Coble, J. R., Virtual Reality Therapy: An Effective Treatment for Psychological Disorders. Virtual Reality in Neuro-Psycho-Physiology, Giuseppe Riva (Ed.), Ios Press: Amsterdam, Netherlands, 1997. PMID: 10175343

[19] Ekman, P., Universals and cultural differences in facial expressions of emotion, Nebraska Symposium on Motivation, 19, p. 207–283, 1971.

[20] **Moise, G.**, Vladoiu, M., Constantinescu, Z., GC-MAS – A Multiagent System for Building Creative Groups Used in Computer Supported Collaborative Learning. In: Jezic, G., Kusek, M., Lovrek, I., J. Howlett, R., Jain, L. (eds) Agent and Multi-Agent Systems: Technologies and Applications. Advances in Intelligent Systems and Computing, vol 296. Springer, Cham. 2014. https://doi.org/10.1007/978-3-319-07650-8_31.

[21] **Moise, G.**, Vladoiu, M., Constantinescu, Z., Building the Most Creative and Innovative Collaborative Groups Using Bayes Classifiers. In: Panetto, H., et al. On the Move to Meaningful Internet Systems. OTM 2017 Conferences. International Conference on Cooperative Information Systems (CoopIS) 2017, Lecture Notes in Computer Science(), vol 10573. Springer, Cham. https://doi.org/10.1007/978-3-319-69462-7_17.

[22] Vladoiu, M., **Moise, G.**, & Constantinescu, Z., Towards Building Creative Collaborative Learning Groups Using Reinforcement Learning. In B. Andersson, B. Johansson, S. Carlsson, C. Barry, M. Lang, H. Linger, & C. Schneider (Eds.), Designing Digitalization (ISD2018 Proceedings). Lund, Sweden: Lund University. ISBN: 978-91-7753-876-9. 2018. http://aisel.aisnet.org/isd2014/proceedings2018/Education/9

[23] **Moise, G.**, Vladoiu, M., & Constantinescu, Z. (2018). Towards Construction of Creative Collaborative Teams Using Multiagent Systems. In B. Andersson, B. Johansson, S. Carlsson, C. Barry, M. Lang, H. Linger, & C. Schneider (Eds.), Designing Digitalization (ISD2018 Proceedings). Lund, Sweden: Lund University. ISBN: 978-91-7753-876-9. 2018. http://aisel.aisnet.org/isd2014/proceedings2018/Education/10.

[24] **Moise, G.**, Nicoară, E., S., Chapter 4 - Ethical aspects of automatic emotion recognition in online learning, Editor(s): Santi Caballé, Joan Casas-Roma, Jordi Conesa, In Intelligent Data-Centric Systems, Ethics in Online AI-based Systems, Academic Press, 2024, Pages 71-95, ISBN 9780443188510, https://doi.org/10.1016/B978-0-443-18851-0.00003-2.

[25] **Moise, G.**, Dragomir, E.G., Șchiopu, D. et al. Towards Integrating Automatic Emotion Recognition in Education: A Deep Learning Model Based on 5 EEG Channels. Int J Comput Intell Syst 17, 230, 2024. https://doi.org/10.1007/s44196-024-00638-x.

[26] Pekrun, R., Goetz, T., Titz W. & Perry R. P., Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research, Educational Psychologist, 37:2, 91-105, 2002. DOI: 10.1207/S15326985EP3702_4

[27] Yadegaridehkordi E., Noor N.F.B.M., Ayub M.N.B., Affal H.B. & Hussin N.B., Affective computing in education: A systematic review and future research, Computers & Education, 2019. doi: https://doi.org/10.1016/j.compedu.2019.103649.

[28] D'Mello, S., A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. Journal of Educational Psychology, 105(4), 1082–1099. 2013. https://doi.org/10.1037/a0032674

[29] Leslie, D., Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute, 2019. https://doi.org/10.5281/zenodo.3240529.

[30] Mitruț, O., Moise, G., Moldoveanu, A. et al. Clarity in complexity: how aggregating explanations resolves the disagreement problem. Artif Intell Rev 57, 338, 2024. https://doi.org/10.1007/s10462-024-10952-7

[31] Shin, D., Park, Y. J., Role of fairness, accountability, and transparency in algorithmic affordance, Computers in Human Behavior, Volume 98, 2019, Pages 277-284, ISSN 0747-5632, https://doi.org/10.1016/j.chb.2019.04.019.

[32] Brughmans, D., Melis L, Martens D., Disagreement amongst counterfactual explanations: How transparency can be deceptive. 2023. arXiv [csAI]. http://arxiv.org/abs/2304.12667

[33] Krishna, S., Han, T., Gu, A., Jabbari, S., Wu, Z.S., Lakkaraju, H., The disagreement problem in explainable machine learning: A practitioner's perspective. Research Square. 2023. http://arxiv.org/abs/2202.01602

[34] Müller, S., Toborek, V., Beckh, K., Jakobs, M., Bauckhage, C., Welke, P., An Empirical Evaluation of the Rashomon Effect in Explainable Machine Learning. In: Koutra, D., Plant, C., Gomez Rodriguez, M., Baralis, E., Bonchi, F. (eds) Machine Learning and Knowledge Discovery in Databases: Research Track. ECML PKDD 2023. Lecture Notes in Computer Science(), vol 14171. Springer, Cham. 2023. https://doi.org/10.1007/978-3-031-43418-1_28.

[35] Pirie, C., Wiratunga, N., Wijekoon, A., Moreno-Garcia, C.F., AGREE: a feature attribution aggregation framework to address explainer disagreements with alignment metrics. In Proceedings of the Workshops at the 31st International Conference on Case-Based Reasoning (ICCBR-WS 2023), pp184–199. CEUR. 2023. https://ceur-ws.org/Vol-3438/paper_14.pdf Accessed April 2025.

[36] Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., Johannes, R.S., Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus. Proc Annu Symp Comput Appl Med Care. 1988 Nov 9:261–5. PMCID: PMC2245318.

[37] Ramana, B., Venkateswarlu, N., ILPD (Indian Liver Patient Dataset). UCI Machine Learning Repository. 2012. https://doi.org/10.24432/C5D02C. Accessed January 2024.

[38] Hepatitis, 1988, UCI Machine Learning Repository. https://doi.org/10.24432/C5Q59J. Accessed January 2024.

[39] Campos, D., Bernardes, J., Cardiotocography. 2010. UCI Machine Learning Repository. https://doi.org/10.24432/C51S4N. Accessed January 2024.

[40] Nash, W., Sellers, T., Talbot, S., Cawthorn, A., & Ford, W., Abalone [Dataset]. 1994. UCI Machine Learning Repository. https://doi.org/10.24432/C55C7W.

[41] Kadiwal, A., Water quality dataset. 2021. https://www.kaggle.com/datasets/adityakadiwal/water-potability. Accessed January 2024

[42] Picard, R., Affective Computing, Cambridge: MIT Press, 1997.

[43] Calvo, R., D'Mello, S., Gratch, J., Kappas, A., Introduction to Affective Computing, chapter in The Oxford Handbook of Affective Computing, Oxford University Press, 2015.

[44] Shouse, E., Feeling, Emotion, Affect. M/C Journal, 8(6). 2005. https://doi.org/10.5204/mcj.2443.

[45] Niven, K., Affect. In: Gellman, M.D., Turner, J.R. (eds) Encyclopedia of Behavioral Medicine. Springer, New York, NY. 2013. https://doi.org/10.1007/978-1-4419-1005-9_1088.

[46] Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., Gao, S., Sun, Y., Ge, W., Zhang, W., Zhang, W., A systematic review on affective computing: emotion models, databases, and recent advances, Information Fusion, Volumes 83–84, 2022, Pages 19-52, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2022.03.009.

[47] Pekrun, R., Emotions and learning. Educational Practices Series, 24. https://www.iaoed.org/downloads/edu-practices_24_eng.pdf, 2014. Accesed January 2024.

[48] Frenzel, A.C., Daniels L. & Irena Burić: Teacher emotions in the classroom and their implications for students, Educational Psychologist, 56:4, 250-264, 2021. https://doi.org/10.1080/00461520.2021.1985501.

[49] Cowie, R., Ethical Issues in Affective Computing, chapter in The Oxford Handbook of Affective Computing, Oxford University Press, 2015

[50] Mandler, G., The generation of emotion: A psychological theory, in Emotion: Theory, research, and experience. Vol. 1: Theories of emotion, New York, Academic Press, 1980.

[52] Tzirakis, P., Trigeorgis, G., Nicolaou, M.A., Schuller, B.W., Zafeiriou, S.m End-to-end multimodal emotion recognition using deep neural networks, IEEE J. Sel. Top. Signal Process. 11, 2017, 1301–1309. https://doi.org/10.1109/ JSTSP.2017.2764438.

[53] Ezzameli, K., Mahersia, H., Emotion recognition from unimodal to multimodal analysis: A review, Information Fusion, Volume 99, 2023, 101847, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2023.101847.

[54] LeDoux, J., Rethinking the emotional brain, Neuron, 73(4), Erratum in: Neuron. 2012 Mar 8;73(5):1052. PMID: 22365542; PMCID: PMC3625946., pp. 653-76, 2012.

[55] Izard, C., Human Emotions, Springer, Boston, MA, 1977, pp. 260-80.

[56] Ekman, P.: Basic emotions. In Handbook of Cognition and Emotion; Dalgleish, T., Power, M., Eds.; John Wiley&Sons Ltd.: Hoboken, NJ, USA, 1999.

[57] Plutchik, R. Emotion: A Psychoevolutionary Synthesis; Harper & Row: New York, NY, USA, 1980.

[58] Cohen, M. A., Against Basic Emotions, and Toward a Comprehensive Theory, Journal of Mind and Behavior, 26(4), pp. 229-254, 2005. http://www.jstor.org/stable/43854066. Accessed January 2025.

[59] Russell, J., A circumplex model of affect, J. Personal. Soc. Psychol, 39, p. 1161–1178, 1980.

[60] Mehrabian A., and Russel, J.A., An Approach to Environmental Psychology, The Massachusetts Institute of Technology: Cambridge, MA, USA, 1974.

[61] Russel J., and Mehrabian, A., Evidence for a three-factor theory of emotions, J. Res. Personal, 11, p. 273–294, 1977. https://doi.org/10.1016/0092-6566(77)90037-X

[62] Mehrabian, A., Pleasure-Arousal-Dominance: A general framework for describing and measuring individual differences in temperament, Curr. Psychol., 14, pp. 261-292, 1996. https://doi.org/10.1007/BF02686918.

[63] Mehrabian, A., Framework for a comprehensive description and measurement of emotional states, Genet. Soc. Gen. Psychol. Monogr, 121, pp. 339-361, 1995. PMID: 7557355.

[64] Buechel, S. & Hahn, U., Emotion analysis as a regression problem - dimensional models and their implications on emotion representation and metrical evaluation, in ECAI'16: Proceedings of the Twenty-second European Conference on Artificial Intelligence, 2016. https://doi.org/10.3233/978-1-61499-672-9-1114.

[65] Woaswi, W. & Hanif, M. & Mohamed, S. & Hamzah, N. & Rizman, Z., Human Emotion Detection via Brain Waves Study by Using Electroencephalogram (EEG). International Journal on Advanced Science, Engineering and Information Technology. 6. 1005. 2016. 10.18517/ijaseit.6.6.1072.

[66] Liu, X., Li, T., Tang, C., Xu, T., Chen, P., Bezerianos A., Wang, H.: Emotion recognition and dynamic functional connectivity analysis based on EEG. IEEE Access 7, 143293–143302, 2019. https://doi.org/10.1109/ACCESS.2019.2945059

[67] Islam, Md. R., Islam, Md.M., Rahman, M.M., Mondal, C., Singha, S.K., Ahmad, M., Awal, A., Islam, Md.S., Moni, M.A.: EEG Channel Correlation Based Model for Emotion Recognition. Computers in Biology and Medicine. 136, 104757, 2021. https://doi.org/10.1016/j.compbiomed.2021.104757.

[68] Soleymani, M., Lichtenauer, J., Pun, T. and Pantic, M., A Multimodal Database for Affect Recognition and Implicit Tagging, in IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 42-55, Jan.-March 2012, doi: 10.1109/T-AFFC.2011.25.

[69] Goshvarpour, A.; Abbasi, A. Dynamical analysis of emotional states from electroencephalogram signals. Biomed. Eng. Appl. Basis Commun. 2016, 28, 1650015. https://doi.org/10.4015/S1016237216500150.

[70] Chai, X.; Wang, Q.S.; Zhao, Y.P.; Liu, X.; Bai, O.; Li, Y.Q. Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition. Comput. Biol. Med. 2016, 79, 205–214. https://doi.org/10.1016/j.compbiomed.2016.10.019,

[71] Zhang, L., Wu, Y., Zhang, L., Wang, Y., Li, M., Synthesis and characterization of mesoporous alumina with high specific area via coprecipitation method, Vacuum, Volume 133, 2016, Pages 1-6, ISSN 0042-207X, https://doi.org/10.1016/j.vacuum.2016.08.005.

[72] Yadava, M., Kumar, P., Saini, R. et al. Analysis of EEG signals and its application to neuromarketing. Multimed Tools Appl 76, 19087–19111 , 2017. https://doi.org/10.1007/s11042-017-4580-6

[73] Zhang, J., Patel, V.L., Johnson, K. A., Malin, J., Smith, J.W., Designing Human-Centered Distributed Information Systems. IEEE Intelligent Systems 17(5), 42-47, 2002. doi: 10.1109/MIS.2002.1039831.

[74] van der Bijl-Brouwer, M., Dorst, K., Advancing the Strategic Impact of Human-Centered Design. 0142-694X Design Studies, 53, 1-23, 2017. https://doi.org/10.1016/j.destud.2017.06.003.

[75] International Classification of Diseases ICD (2025), ICD-11 for Mortality and Morbidity Statistics, URL: https://icd.who.int/browse/2025-01/mms/en, Accessed March 2025.

[76] Institute for Health Metric and Evaluation, https://www.healthdata.org/research-analysis/health-risks-issues/mental-health, Accessed January 2025

[77] Garcia-Palacios, H. G., Hoffman, S., Kwong See, A., Botella, C., Redefining Therapeutic Success with Virtual Reality Exposure Therapy". CyberPsychology & Behavior, 4(3), 341–348, 2001. http://online.liebertpub.com/doi/abs/10.1089/109493101300210231

[78] Arikan, K.; Boutros, N.N.; Bozhuyuk, E.; Poyraz, B.C.; Savrun, B.M.; Bayar, R.; Gunduz, A.; Karaali-Savrun, F.; Yaman, M. EEG Correlates of Startle Reflex with Reactivity to Eye Opening in Psychiatric Disorders: Preliminary Results. Clin. EEG Neurosci. 2006, 37, 230–234. DOI: 10.1177/155005940603700313

[79] Kometer, H.; Luedtke, S.; Stanuch, K.; Walczuk, S.; Wettstein, J. The Effects Virtual Reality Has on Physiological Responses as Compared to Two-Dimensional Video. J. Adv. Stud. Sci., 1, pp. 1-21, 2010.

[80] Koestler, A.: The Ghost in The Machine. Arkana, New York, 1967.

[81] Garcia-Herreros, E., Christensen, J., Prado, J.M., Tamura, S.: IMS - holonic manufacturing systems: System components of autonomous modules and their distributed control. Technical report, HMS Consortium, 1994.

[82] **Moise, G.**, Moise, P.G., Moise, P.S.: Towards holons-based architecture for medical systems. In: Proceedings of 2018 ACM/IEEE International Workshop on Software Engineering in Healthcare Systems SEHS 2018, Gothenburg, Sweden, pp. 26–30, 2018. https://doi.org/10.1145/3194696.3194702.

[83] Christensen, J.H., Holonic manufacturing systems: initial architecture and standard directions. In: Proceedings of the First European Conference on Holonic Manufacturing Systems, Hannover,1994.

[84] Laal, M., Laal, M., Collaborative learning: what is it?, Procedia - Social and Behavioral Sciences, Volume 31, 2012, Pages 491-495, ISSN 1877-0428, https://doi.org/10.1016/j.sbspro.2011.12.092.

[85] Stahl, G., Koschmann, T., Suthers, D., Computer-Supported Collaborative learning: An historical perspective. In Sawyer R. K. (ed.), Cambridge handbook of the learning sciences, pp. 409-426, Cambridge University Press, Cambridge, UK, (2006)

[86] Sternberg, R.J., Lubart, T.I. , Kaufman, J.C., Pretz, J. E., Creativity. In Holyoak K. J., Morrison, R. G., The Cambridge Handbook of Thinking and Reasoning, pp 351-369, Cambridge University Press, New York, 2005.

[87] Sternberg, R. J., Lubart, T. I., An Investment Theory of Creativity and its Development, Human Development. 34(1), pp. 1–31, 1991.

[88] Sternberg, R. J., The Assessment of Creativity: An Investment-Based Approach, Creativity Research Journal. 24(1), pp. 3–12, 2012. https://doi.org/10.1080/10400419.2012.652925

[89] Gorny, E. (ed.): Group creativity. Dictionary of Creativity: Terms, Concepts, Theories &Findings in Creativity Research, 2007. http://creativity.netslova.ru/Group_creativity.html. Accessed April 2025.

[90] Amabile, T. M., A Model of Creativity and Innovation in Organizations. Research in organizational behavior. In Staw, B. M., Cummings, L. L. (eds.), Research in organizational behavior, 10, 123-167. Greenwich, CT: JAI Press, 1988.

[91] Amabile, T. M., Social Psychology of Creativity: A Componential Conceptualization. Journal of Personality and Social Psychology 45(2), 357-377, 1983. https://doi.org/10.1037/0022-3514.45.2.357.

[92] Amabile, T. M.: Componential Theory of Creativity. In: Kessler, E. H., (ed.). Encyclopedia of Management Theory. pp. 135-140. SAGE Publications Inc. (2013)

[93] Amabile, T. M., Componential Theory of Creativity. Working paper, https://www.hbs.edu/ris/Publication%20Files/12-096.pdf, Accessed April 2025.

[94] Woodman, R.W., Schoenfeldt, L. F., Individual Differences in Creativity: An Interactionist Perspective. In Glover, J.A., Ronning, R. R., Reynolds, C. R. (eds.), Handbook of creativity, pp. 77-92, Plenum Press, New York, 1989.

[95] Woodman, R.W., Schoenfeldt, L. F., An Interactionist Model of Creative Behaviour. Journal of Creative Behavior. 24, pp. 279-290, 1990. **https://doi.org/10.1002/j.2162-6057.1990.tb00549.x**.

[96] Woodman, R.W., Sawyer, J. E., Griffin, R. W., Toward a Theory of Organizational Creativity, Academy of Management Review. 18, No. 2, pp. 293-321, 1993.

[97] Watkins, C.: Learning from Delayed Rewards, PhD Thesis,University of Cambridge, England, (1989), http://www.cs.rhul.ac.uk/home/chrisw/thesis.html, Accessed January 2025.

[98] Choi, D. Y., Lee, K. C. and Seo, Y. W.: Scenario-Based Management of Team Creativity in Sensitivity Contexts: An Approach with a General Bayesian Network. In: Lee, K. C. (Ed.), Digital Creativity Individuals, Groups, and Organizations, 2013. https://doi.org/10.1007/978-1-4614-5749-7_7.

[99] Paulus, P. B., & Dzindolet, M. T., Social influence, creativity and innovation. Social Influence (3), 228–247. 2008. https://doi.org/10.1080/15534510802341082.

[100] Reiter-Palmon, R., Wigert, B., and de Vreede, T.: Team Creativity and Innovation: The Ef-fect of Group Composition, Social Processes, and Cognition. In Handbook of Organizational Creativity, Elsevier Science & Technology, 2011.

[101] Hascher, T., Learning and Emotion: Perspectives for Theory and Research. European Educational Research Journal, 9(1), 13–28. 2010. https://doi.org/10.2304/eerj.2010.9.1.13

[102] Pekrun, R., Progress and open problems in educational emotion research. Learning and Instruction, 15(5), 497–506. 2005. https://doi.org/10.1016/j.learninstruc.2005.07.014..

[103] Um, E. "R.", Plass, J. L., Hayward, E. O., & Homer, B. D., Emotional design in multimedia learning. Journal of Educational Psychology, 104(2), 485–498. 2012. https://doi.org/10.1037/a0026609

[104] Boekaerts, M., & Pekrun, R., Emotions and emotion regulation in academic settings. In L. Corno & E. M. Anderman (Eds.), Handbook of educational psychology (3rd ed., pp. 76–90). Routledge/Taylor & Francis Group, 2016.

[105] Hascher, T. & Edlinger, H. Positive Emotionen und Wohlbefinden in der Schule – ein Rubberlike über Forschungszugänge und Erkenntnisse [Positive emotions and well-being in school – an overview of methods and results], Psychologie in Erziehung und Unterricht, 56, 105-122, 2009.

[106] D'Mello, S., Lehman, B., Pekrun, R., Graesser, A., Confusion can be beneficial for learning, Learning and Instruction, Volume 29, 2014, Pages 153-170, ISSN 0959-4752, https://doi.org/10.1016/j.learninstruc.2012.05.003.

[107] D'Mello, S. and Graesser, A. : AutoTutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. ACM Trans. Interact. Intell. Syst., vol. 2, no. 4, Article 23, 1-39 (2012). https://doi.org/10.1145/2395123.2395128.

[108] Sidney D'Mello, Andrew Olney, Claire Williams, Patrick Hays, Gaze tutor: A gaze-reactive intelligent tutoring system, International Journal of Human-Computer Studies,

Volume 70, Issue 5, 2012, Pages 377-398, ISSN 1071-5819, https://doi.org/10.1016/j.ijhcs.2012.01.004.

[109] Khediri, N., Ben Ammar, M. & Kherallah, M. A Real-time Multimodal Intelligent Tutoring Emotion Recognition System (MITERS). Multimed Tools Appl 83, 57759–57783, 2024. https://doi.org/10.1007/s11042-023-16424-4

[110] Kort, B., Reilly, R. and Picard, R. W.: An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion. In: Proceedings IEEE International Conference on Advanced Learning Technologies, Madison, WI, USA, 43-46, 2001. https://doi.org/10.1109/ICALT.2001.943850

[111] Shen, L., Wang, M., & Shen, R., Affective e-Learning: Using "Emotional" Data to Improve Learning in Pervasive Learning Environment. Educational Technology & Society, vol. 12, no. 2, 176–189, 2009.

[112] Mohamad Nezami, O., Dras, M., Hamey, L., Richards, D., Wan, S., Paris, C.: Automatic Recognition of Student Engagement Using Deep Learning and Facial Expression. In: Brefeld, U., Fromont, E., Hotho, A., Knobbe, A., Maathuis, M., Robardet, C. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2019. Lecture Notes in Computer Science, vol. 11908, Springer, Cham, 2020. https://doi.org/10.1007/978-3-030-46133-1_17.

[113] Akter, S., Prodhan, R.A., Pias, T.S., Eisenberg, D., Fresneda Fernandez, J.: M1M2: Deep-Learning-Based Real-Time Emotion Recognition from Neural Activity. Sensors. 22, 8467, 2022. https://doi.org/10.3390/s22218467.

[113] Alyuz, N., Okur, E., Oktay, E., Genc, U., Aslan, S., Mete, S.E., Arnrich, B., and Esme, A.A.: Semi-supervised model personalization for improved detection of learner's emotional engagement. In: Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16). Association for Computing Machinery, New York, NY, USA, 100–107, 2016. https://doi.org/10.1145/2993148.2993166.

[114] . Topic, A., Russo, M., Stella, M., Saric, M.: Emotion Recognition Using a Reduced Set of EEG Channels Based on Holographic Feature Maps. Sensors 22, 3248, 2022. https://doi.org/10.3390/s22093248.

[115] Mohammad, S.M., Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis, Computational Linguistics, 48(2):239–278, 2002. https://doi.org/10.1162/coli_a_00433.

[116] CDDO (Central Digital & Data Office). (2020). Guidance Data Framework, https://www.gov.uk/government/publications/data-ethics-framework, Accessed April 2025.

[117] Regulation (EU) 2016 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 199/1.

[118] Miller, T., Explanation in artificial intelligence: Insights from the social sciences, Artificial Intelligence, Volume 267, 2019, Pages 1-38, ISSN 0004-3702, https://doi.org/10.1016/j.artint.2018.07.007.

[119] Biran O., & Cotton, C., Explanation and Justification in Machine Learning: A Survey, IJCAI-17 workshop on explainable AI (XAI), 2017.

[120] Roscher, R., Bohn, B., Duarte M. F. and Garcke, J., Explainable Machine Learning for Scientific Insights and Discoveries, in IEEE Access, vol. 8, pp. 42200-42216, 2020, doi: 10.1109/ACCESS.2020.2976199.

[121] Adadi A., M. Berrada, M., Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), in IEEE Access, vol. 6, pp. 52138-52160, 2018, doi: 10.1109/ACCESS.2018.2870052.

[122] Christoph Molnar, M., Interpretable Machine Learning. A Guide for Making Black Box Models Explainable, 2025, https://christophm.github.io/interpretable-ml-book/, Accessed April 2025.

[123] Juliussen, B. A., The Right to an Explanation Under the GDPR and the AI Act. In MultiMedia Modeling: 31st International Conference on Multimedia Modeling, MMM 2025, Nara, Japan, January 8–10, 2025, Proceedings, Part IV. Springer-Verlag, Berlin, Heidelberg, 184–197. 2025. https://doi.org/10.1007/978-981-96-2071-5_14.

[124] Vainio-Pekka, H., Ori-Otse Agbese, M., Jantunen, M., Vakkuri, V., Mikkonen, T., Rousi, R., and Abrahamsson. P., The Role of Explainable AI in the Research Field of AI Ethics. ACM Trans. Interact. Intell. Syst. 13, 4, Article 26, 2023, 39 pages. https://doi.org/10.1145/3599974.

[125] Ribeiro, M.T., Singh, S., Guestrin, C.: "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), Association for Computing Machinery, New York, NY, USA, 1135–1144, 2016. https://doi.org/10.1145/2939672.2939778

[126] Chakraborti, S., Beresi, U., Wiratunga, N., Massie, S., Lothian, R., Watt, S., A Simple Approach towards Visualizing and Evaluating Complexity of Textual Case Bases. In: Proc. of the ICCBR 2007 Workshops. 2007.

[127] Holzinger, A., Interactive machine learning for health informatics: when do we need the human-in-the-loop. Brain Informatics, 2, 119–131, 2016. DOI 10.1007/s40708-016-0042-6

[128] Gough, H. G.: A Creative Personality Scale for the Adjective Check List. Journal of Personality and Social Psychology (37), 1398-1405, 1979. https://doi.org/10.1037/0022-3514.37.8.1398.

[129] Pintrich, P. R., Smith, D. A. F., Garcia, T., McKeachie, W. J., Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (Mslq). In Educational and Psychological Measurement, 53(3), 801-813, 1993. https://doi.org/10.1177/0013164493053003024.

[130] Pilz, K. F., Mahmood, Y., Heim, L., AI's Power Requirements Under Exponential Growth Extrapolating AI Data Center Power Demand and Assessing Its Potential Impact on U.S. Competitiveness, 2025, https://www.rand.org/pubs/research_reports/RRA3572-1.html. Accessed April 2025.

[131] Bashir, N., Donti, P., Cuff, J., Sroka, S., Ilic, M., Sze, V., Delimitrou, C., & Olivetti, E., The Climate and Sustainability Implications of Generative AI. *An MIT Exploration of Generative AI*. 2024. https://doi.org/10.21428/e4baedd9.9070dfe7.

[132] Eshraghian J. K., et al., "Training Spikings Neural Networks Using Lessons From Deep Learning," in Proceedings of the IEEE, vol. 111, no. 9, pp. 1016-1054, Sept. 2023, doi: 10.1109/JPROC.2023.3308088.

[133] Gewaltig, MO., Spiking Network Models and Theory: Overview. In: Jaeger, D., Jung, R. (eds) Encyclopedia of Computational Neuroscience. Springer, New York, NY. 2022. https://doi.org/10.1007/978-1-0716-1006-0_792

[134] Lapicquem L., Recherches quantitatives sur l'excitation electrique des nerfs traitee comme une polarization. J Physiol Pathol Gen 9:620–635. 1907.

[135] Gerstner, W., Hemmen, J.V., Cowan, J., What matters in neuronal locking? Neural Comput 8:1653–1676, 1996.

[136] Gerstner, W., Kistler, W.M., Spiking neuron models: single neurons, populations, plasticity. Cambridge University Press, Cambridge, 2002.

[137] Mollahosseini, A., Hasani B., and Mahoor, M. H., AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild, in *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18-31, 1 Jan.-March 2019, doi: 10.1109/TAFFC.2017.2740923.

[138] Nguyen, E., Nauta, M., Englebienne G., and Seifert, C., Feature Attribution Explanations for Spiking Neural Networks, 2023 IEEE 5th International Conference on Cognitive Machine

Intelligence (CogMI), Atlanta, GA, USA, 2023, pp. 59-68, 2023. doi: 10.1109/CogMI58952.2023.00018.

[139] Bitar, A., Rosales, R., Paulitsch, M., Gradient-based feature-attribution explainability methods for spiking neural networks, Front. Neurosci., 27 September 2023, Sec. Neuromorphic Engineering, Volume 17 - 2023, https://doi.org/10.3389/fnins.2023.1153999.

[140] Moon, J., Pedretti, G., Bruel, P., Serebryakov, S., Eldash, O., Buonanno, L., Graves, C. E., Faraboschi, P., and Ignowski. J., CAMSHAP: Accelerating Machine Learning Model Explainability with Analog CAM. In Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design (ICCAD '24). Association for Computing Machinery, New York, NY, USA, Article 86, 1–9. 2025. https://doi.org/10.1145/3676536.3676696.

[141] Zhang, B., Anderljung, M., Kahn, L., Dreksler, N., Horowitz, M.C. and Dafoe, A., Ethics and Governance of Artificial Intelligence: Evidence from a Survey of Machine Learning Researchers. J. Artif. Int. Res. 71, 2021, 591–666. https://doi.org/10.1613/jair.1.12895.

[142] Hanna, M. G., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J., Deebajah, M., Rashidi, H. H., Ethical and Bias Considerations in Artificial Intelligence/Machine Learning, Modern Pathology, Volume 38, Issue 3, 2025, 100686,ISSN 0893-3952, https://doi.org/10.1016/j.modpat.2024.100686.