MULTIMODAL BIRLASHTIRISH MODELI YORDAMIDA YUZ TASVIRI VA NUTQ BELGILARI BO‘YICHA SHAXSNI IDENTIFIKATSIYALASH SAMARADORLIGINI OSHIRISH

Ergasheva Durdona Sherali qizi
Namangan viloyati Uchqo‘rg‘on tumani
20-sonli umumiy o‘rta ta’lim maktabi
matematika fani o‘qituvchisi
E-mail: dergasheva209@gmail.com
https://doi.org/10.5281/zenodo.23274051

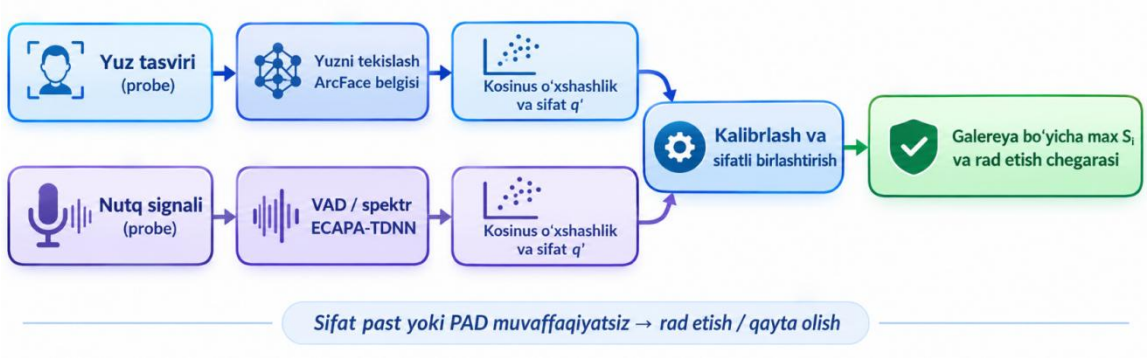
Biometrik identifikatsiya kuzatuv namunasi galereyada ro‘yxatdan o‘tgan qaysi shaxsga tegishli ekanini aniqlashga qaratilgan 1:N qidiruvdir; verifikatsiyada esa oldindan bildirilgan bitta shaxs bilan 1:1 solishtirish bajariladi. Shu sababli 1:N uchun galereya hajmi va “galereyada yo‘q” holatini hisobga oladigan FPIR hamda FNIR ko‘rsatkichlari kerak [1].

Yuz tasviri yorug‘lik, burilish, berkilish va kamera sifatiga, nutq esa shovqin, kanal va davomiylikka sezgir. ArcFace yuz vektorlarini burchakli marginli yo‘qotish orqali, ECAPA-TDNN esa nutqdan so‘zlovchiga xos vektorni ajratadi [1, 2]. FaceNet va GMM ballarini birlashtirgan avvalgi ish ikki kanalni birga qo‘llash mumkinligini ko‘rsatgan, biroq natijalarini boshqa ma’lumot va protokolga ko‘chirib bo‘lmaydi [3].

Tadqiqotning maqsadi yuz va nutq kanallarining o‘zgaruvchan sifatini hisobga oladigan multimodal birlashtirish modeli orqali 1:N identifikatsiya samaradorligini oshirish imkoniyatini baholashdir. Hissa yangi kodlagich yoki rekord aniqlik emas: kalibrlangan ballarni sifatga mos vaznlash, ochiq to‘plamdagi qaror va rad etish qoidasi hamda yaxshilanishni mustaqil testda o‘lchash protokoli yagona modelda ifodalanadi. Kanal xatolari mustaqil emas (sifatsiz telefon kamerasi va mikrofoni ikkalasini birga yomonlashtiradi), shuning uchun ikki kanal qaysi sharoit va chegarada, qanday xatolik narxi evaziga bittadan yaxshiroq ekani aniqlanadi. Gipotezalar: H1 – bir xil FPIRda, ayniqsa bitta kanal sifati pasayganda, sifatga mos vaznlash teng vaznli birlashtirishdan kichikroq FNIR beradi; H2 – sifat bilan bog‘liq rad etish yaroqsiz yozuvlar yoki hujumlarning avtomatik qabul qilinishini kamaytiradi; H3 – foyda barcha demografik va qurilma guruhlarida bir xil bo‘lmasligi mumkin. H1–H3 tekshirilmaguncha natija sifatida ifodalanmaydi.

Ma’lumot birligi va ajratish qoidasida asosiy birlik – rozilik bilan olingan bir shaxsga tegishli yuz tasviri va nutq yozuvi juftligi (x^f, x^v, y). Galereya va so‘rov (probe) yozuvlari turli sessiyalardan olinadi; bir sessiyadagi takroriy kadrlar ikki tomonga tarqatilmaydi. Mashq, kalibrlash va yakuniy testda shaxslar kesishmaydi; ochiq to‘plam uchun galereyaga kirmagan shaxslarning probelari ham ajratiladi. VoxCeleb2 bir milliondan ortiq nutq namunasini [4], VGGFace2 esa 9131 shaxsning 3,31 million tasvirini [5] qamraydi, ammo ular bilan tajriba o‘tkazilgani da’vo qilinmaydi va ular yozuv darajasida tayyor audio–yuz juftligi emas: haqiqiy multimodal sinov uchun shaxs IDsi, sessiya va vaqt belgisi mos, foydalanish huquqi tekshirilgan juftliklar kerak. Ishonch oralig‘ida qayta tanlash birligi kadr emas, subyekt bo‘ladi. Tayyor vaznlarning o‘qitish ma’lumotlari bilan kesishish ehtimoli ochiq aytiladi.

Belgilarni ajratish va sifat nazoratida tasvirda yuz topilib tekislanadi; ArcFace birlik uzunlikka keltirilgan $e^f \in R^d$ vektorini hosil qiladi [1]. Audioda ovoz faolligi aniqlanib, ECAPA-TDNN $e^v \in R^k$ so‘zlovchi vektorini beradi [2]. Vektorlar turli fazoda bo‘lgani uchun elementma-element qo‘shilmaydi. Har bir probe uchun q^f va $q^v \in [0,1]$ sifat ko‘rsatkichi alohida hisoblanadi (tasvirda fokus, aniqlanish ishonchi, burilish; audioda samarali nutq davomiyligi, shovqin); mezonlar rivojlantirish to‘plamida oldindan belgilanadi va testda moslanmaydi. PAD (taqdimot hujumini aniqlash) natijasi sifat bilan bir tushuncha emas: hujumga gumon qilingan namunani yaxshi fokus yoki baland ovoz “qutqarmaydi” (1-rasm) [7].



1-rasm. Taklif etilgan yuz va nutq belgilarini ballar darajasida birlashtirish sxemasi.

Multimodal birlashtirish formulasi va qaror qoidasi har bir galereya shaxsi j uchun yuz va nutq vektorlarining mos etalon bilan kosinus o‘xshashligi s^f_j va s^v_j hisoblanadi. Ballar faqat rivojlantirish to‘plamida baholangan o‘rtacha μ^m va standart og‘ish $\sigma^m > 0$ orqali standartlashtiriladi (1), bunda $m \in \{f, v\}$; bu ehtimollik kalibrlashi emas.

$$s^m_j = \cos(e^m, g^m_j), \quad z^m_j = (s^m_j - \mu^m) / \sigma^m \qquad (1)$$

Signal mavjud, sifati $q^m \geq q^m_{min}$ va PAD tekshiruvidan o‘tgan bo‘lsa $a^m = 1$, aks holda $a^m = 0$. Kanal vazni (2), birlashtirilgan ball (3) bo‘yicha olinadi:

$$w^m = a^m q^m / (a^f q^f + a^v q^v) \qquad (2)$$

$$S_j = w^f z^f_j + w^v z^v_j \qquad (3)$$

Ikkala kanal faol bo‘lmasa qaror chiqarilmaydi; faqat bitta kanal qolsa natija “bir kanalli” deb belgilanadi va shu rejim uchun rivojlantirishda alohida chegara tanlanadi. 1:N qidiruvda eng katta ballli j^* nomzod tanlanadi: $S_{j^*} \geq \tau$ bo‘lsa nomzod qaytariladi, aks holda “galereyada aniqlanmadi” javobi beriladi. τ rivojlantirish to‘plamida maqsadli FPIR uchun bir marta tanlanadi, galereya hajmi o‘zgarsa qayta tekshiriladi. q sensor signalining yaroqliligini ifodalaydi, identifikatsiya to‘g‘riligi ehtimoli emas. Yuz va ovoz turli nomzodni yuqori qo‘ysa, “kelishmovchilik” bayrog‘i qarorni inson tekshiruviga yuboradi; bu bayroq mustaqil xavfsizlik isboti emas.

Baholash rejasi aynan bir xil galereya va probe tarkibida 1-jadvaldagi variantlar solishtiriladi. Yopiq to‘plamda Rank-1 va Rank-5, ochiq to‘plamda oldindan belgilangan FPIR = α nuqtasidagi FNIR olinadi [6]; teng vaznli (EW) usulga nisbatan yaxshilanish (4) bilan o‘lchanadi (ΔFNIR > 0 – yaxshilanish):

$$\Delta FNIR = FNIR_{EW}(\alpha) - FNIR_Q(\alpha) \qquad (4)$$

Rank-1 galereyada bor probe uchun to‘g‘ri shaxs birinchi o‘ringa chiqqan ulushdir; ochiq to‘plamda nomzod balli τ dan ham o‘tishi kerak, shu sabab reyting sifati va FNIR @ FPIR alohida beriladi. 1:1 dagi FMR yoki EER 1:N natijani almashtirmaydi: galereya kattalashganda begona probe uchun kamida bitta nomzod chegaradan oshish ehtimoli o‘zgaradi [6]. Rad etilgan yozuvlar ulushi va hujumlarni qabul qilish ko‘rsatkichi ISO/IEC 30107-3:2023 doirasida hujum turi bo‘yicha alohida beriladi [7]. Sinovlar sifat va hujum sharoitlari bo‘yicha ajratiladi (toza, past yorug‘lik, shovqin, qisqa nutq, bir kanal yo‘qligi); har bir guruh uchun namuna soni, subyektlar bo‘yicha qayta tanlangan 95% ishonch oralig‘i va juftlangan bootstrap farqi beriladi, demografik kichik guruhlar xatolari ham baholanadi (1-jadval) [8].

1-jadval. Kelgusidagi empirik taqqoslash uchun oldindan belgilangan konfiguratsiyalar

Variant	Faol hisoblash	Asosiy savol
F	Faqat kalibrlangan yuz balli	Yuz yakka qanchalik barqaror?
V	Faqat kalibrlangan ovoz balli	Ovoz yakka qanchalik barqaror?
F + V	Ikki ballning teng vaznli yig‘indisi	Ikki kanalning o‘zi foyda beradimi?
Q-FV	q^f va q^v bo‘yicha vaznli yig‘indi	Sifat ma’lumoti qo‘shimcha foyda beradimi?
Q-FV + PAD	Sifatli yig‘indi va hujumni rad etish	Qabul va hujum xatolari qanday o‘zgaradi?

Maxfiylik va xavfsizlik talablari esa “shaxsga doir ma’lumotlar to‘g‘risida”gi qonun biometrik ma’lumotlarni anatomik va fiziologik xususiyatlarni tavsiflovchi ma’lumotlar deb belgilaydi; shaxsni aniqlash maqsadida ishlov berish rozilikni talab qiladi, qonunda nazarda tutilgan istisnolar bundan mustasno [12]. Qonunning 27^1-moddasi biometrik ma’lumotlarni O‘zbekiston hududida saqlashni talab qiladi [12], Vazirlar Mahkamasining 570-son qarori esa saqlash jismlari va bazadan tashqari saqlash texnologiyalariga talablarni belgilaydi [13]. Xom audio, yuz tasviri va embeddinglar bir-biriga bog‘lanadigan sezgir ma’lumot: eng kam zarur ma’lumot yig‘iladi, shifrlangan saqlanadi, kirish rollar bo‘yicha beriladi, jurnal yuritiladi va chiqqan shaxs yozuvlari o‘chiriladi.

Ushbu metodologik ishda juftlangan yangi shaxslar ma’lumotida yakuniy test o‘tkazilmagan, shuning uchun quyidagilar formuladan kelib chiqadigan analitik xususiyatlar va oldindan belgilangan tekshiruv shartlaridir; (4) dagi yaxshilanishning sonli qiymati yo‘q va samaradorlik oshgani haqida empirik xulosa chiqarilmaydi.

Formuladan keladigan chegaraviy holatlar ikki kanal faol bo‘lsa $w^f + w^v = 1$ va vaznlar manfiy emas. Sifat teng ($q^f = q^v > 0$) bo‘lsa $w^f = w^v = 0{,}5$ va model teng vaznli usul bilan bir xil S_j hamda tartibni qaytaradi; bu dasturiy tekshiruv vazifasini bajaradi. Bir kanal rad etilsa natija bir kanalli rejimga o‘tadi; $a^f = a^v = 0$ bo‘lsa avtomatik identifikatsiya rad etiladi; $S_{j*} < \tau$ bo‘lsa shaxs tayinlanmaydi. Yuz va ovoz turli shaxsni yuqori qo‘ysa, formula ziddiyatni yo‘qotmaydi, shuning uchun modelni “ikki mustaqil tasdiq” deb ta’riflash xato. $a^m = 0$ qoidasi hujumlarni aniqlashning empirik isboti emas: PADni aldab o‘tgan hujum S_j ni oshirishi mumkin, yuqori q esa xatolik pastligini kafolatlamaydi.

Rejalashtirilgan natijalarni taqdim etish shakli empirik bosqichda har bir variant uchun probelar soni, PADdan o‘tmaganlar, Rank-1, belgilangan FPIRdagi FNIR, guruhlar bo‘yicha farq va ishonch oralig‘i bitta jadvalda beriladi; hozir bu kataklarga taxminiy qiymat

qo‘yilmaydi. ROC yoki DET chizig‘i faqat haqiqiy ballar bilan chiziladi; 1-rasmdagi sxema eksperimental grafika emas. 1-jadval natija emas, oldindan qayd etilgan sinov rejasidir.

Ball darajasidagi arxitektura mavjud kodlagichlarni almashtirmasdan va turli o‘lchamli belgilarni qo‘shmasdan integratsiyaga imkon beradi. FaceNet–GMM ishida ham ballar darajasida birlashtirish qo‘llangan [3]; bu maqoladagi qo‘shimcha tamoyil sifatni qaror formulasi va rad etish qoidasi bilan bog‘lashdir. Biroq kanal xatolari korrelyatsiyalashganda, sifat o‘lchovi yomon kalibrlanganda yoki μ va σ domenlararo barqaror bo‘lmaganda yaxshilanish kafolatlanmaydi; formula yoki chegarani yakuniy testga qarab tanlash xolis baholashni buzadi.

Yuz va ovozni surat, video, qayta ijro yoki sun’iy nutq bilan aldash mumkin; ikki ballning yuqoriligi signal tirik shaxsdan kelganini isbotlamaydi. Nol-kuchli begona shaxslardagi FPIR hujum bardoshligini o‘lchamaydi, ISO/IEC 30107-3 esa sensorga ko‘rsatiladigan taqdimot hujumlarini baholaydi va injeksiyaga chidamlilikni o‘z-o‘zidan tasdiqlamaydi [7]. Hozirgi sxema lab sinxronligini tekshirmaydi; bu model tavsifining aniq chegarasidir.

NIST SP 800-63B-4 ovozga asoslangan biometrik solishtirishni man etadi va yuz uchun PADni talab qiladi [9]; shu sababli model NISTga mos kirish autentifikatsiyasi emas, balki ilmiy identifikatsiya tajribasi yoki alohida normativ va etika tekshiruvidan o‘tadigan tadqiqot prototipi sifatida ko‘riladi. Model o‘zbek tilida so‘zlovchilar uchun tayyor aniqlikni isbotlamaydi: til, dialekt, mikrofon va guruhlar bo‘yicha mustaqil tanlamasiz xulosa mahalliy sharoitga ko‘chirilmaydi [10, 11]. Rozilik blankida yuz va nutqning nima uchun ishlatilishi, kim ko‘rishi va qachon o‘chirilishi aniq bo‘lishi kerak; rozilik muddatsiz saqlashga teng emas [12, 13]. Xolislik hisobotida guruhlar bo‘yicha xatolar va rad etish ulushi ishonch oralig‘i bilan beriladi: q bilan vaznlash ayrim guruhlarga yordam berishi ham, ularga ko‘proq rad etish yukini yuklashi ham mumkin, buni faqat sinov hal qiladi [8].

Cheklovlar: juftlangan ma’lumotlar to‘plami va mustaqil test yo‘q; q va q^{m}_{min} sonli optimallashtirilmagan; qurilma, til, aksent va yosh ta’siri o‘lchanmagan. Keyingi bosqichda rozilik asosida yig‘ilgan, turli sessiya va qurilmali korpusda oldindan ro‘yxatlangan protokol bo‘yicha tajriba o‘tkazish, kod va bo‘linish ro‘yxatini e’lon qilish hamda modelni hujum va guruhlar kesimida qayta tekshirish talab etiladi.

Xulosa. Maqolada yuz uchun ArcFace, nutq uchun ECAPA-TDNN belgilariga asoslangan sifatga moslashuvchi multimodal birlashtirish modeli formal tavsiflandi. Ballarni standartlashtirish, sifat vaznlari va ochiq to‘plam qarori raqamlangan formulalarda berildi; samaradorlikni oshirish mezoni bir xil FPIR nuqtasidagi FNIR farqi sifatida belgilandi. Modelning bazaviy usullardan ustunligi hozircha empirik tasdiqlanmagan; xulosa shaxslar bo‘yicha mustaqil, turli sessiyali juftlangan ma’lumotlar sinovidan so‘ng chiqariladi.

Foydalanilgan adabiyotlar ro‘yxati:

1. Deng J., Guo J., Xue N., Zafeiriou S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. CVPR, 2019, 4690–4699 b.
2. Desplanques B., Thienpondt J., Demuynck K. ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in TDNN Based Speaker Verification. Interspeech, 2020, 3830–3834 b.

3. Alharbi B., Alshanbari H. S. Face-voice based multimodal biometric authentication system via FaceNet and GMM. PeerJ Computer Science, 2023, 9:e1468.
4. Chung J. S., Nagrani A., Zisserman A. VoxCeleb2: Deep Speaker Recognition. Interspeech, 2018.
5. Cao Q., Shen L., Xie W., Parkhi O. M., Zisserman A. VGGFace2: A Dataset for Recognising Faces across Pose and Age. FG, 2018. https://www.robots.ox.ac.uk/~vgg/publications/2018/Cao18/
6. NIST. Face Recognition Technology Evaluation: 1:N Identification. https://pages.nist.gov/frvt/html/frvt1N.html
7. ISO/IEC 30107-3:2023. Information technology – Biometric presentation attack detection – Part 3: Testing and reporting. https://www.iso.org/standard/79520.html
8. NIST. Face Recognition Technology Evaluation: Demographic Effects in Face Recognition. https://pages.nist.gov/frvt/html/frvt_demographics.html
9. NIST SP 800-63B-4. Digital Identity Guidelines: Authentication and Authenticator Management. 2025. DOI: 10.6028/NIST.SP.800-63B-4. https://pages.nist.gov/800-63-4/sp800-63b.html
10. Kaxarov Sh. S., Ibragimov A. I. Shaxsni biometrik identifikatsiyalash muammolari. University Research Base, 2024, 383–386 b.
11. Umarov Sh. A., Abduqodirov A. A. Biometrik identifikatsiya uchun neyron tarmoq algoritmlari. Buxoro davlat universiteti ilmiy axboroti, 2024, 8(113), 163 b.
12. O‘zbekiston Respublikasi. “Shaxsga doir ma’lumotlar to‘g‘risida”gi Qonun, O‘RQ-547, 02.07.2019, amaldagi tahrir. https://lex.uz/uz/docs/-4396419
13. O‘zbekiston Respublikasi Vazirlar Mahkamasi. Shaxsga doir ma’lumotlarga ishlov berish sohasidagi ayrim normativ-huquqiy hujjatlarni tasdiqlash to‘g‘risida, 570-son qaror, 05.10.2022. https://lex.uz/docs/-6225462