

**RAQAMLI TEXNOLOGIYALAR VA SUN'YIY INTELLEKTNI
RIVOJLANTIRISH ILMIY-TADQIQOT INSTITUTI
HUZURIDAGI ILMIY DARAJALAR BERUVCHI
DSc.13/30.12.2021.T.142.01 RAQAMLI ILMIY KENGASH**

O'ZBEKISTON MILLIY UNIVERSITETI

MAXAROV QODIRBEK TOLKUNOVICH

**BERILGANLARDA NOANIQLIK BO'LGANDA
QAROR QABUL QILISH TIZIMI**

05.01.11 – Raqamli texnologiyalar va sun'iy intellekt

**TEXNIKA FANLARI BO'YICHA FALSAFA DOKTORI (PhD)
DISSERTATSIYASI AVTOREFERATI**

Toshkent – 2024



**Texnika fanlari bo'yicha falsafa doktori (PhD)
dissertatsiyasi avtoreferati mundarijasi**

**Оглавление автореферата диссертации доктора философии (PhD)
по техническим наукам**

**Contents of dissertation abstract of doctor of philosophy (PhD)
on technical sciences**

Maxarov Qodirbek Tolkunovich

Berilganlarda noaniqlik bo'lganda qaror qabul qilish tizimi 3

Махаров Кодирбек Толкунович

Система принятия решений при наличии неопределённости в данных 21

Makharov Kodirbek Tolkunovich

Decision making system in the presence of uncertainty in data 39

E'lon qilingan ishlar ro'yxati

Список опубликованных работ

List of published works 42



**RAQAMLI TEXNOLOGIYALAR VA SUN'IY INTELLEKTNI
RIVOJLANTIRISH ILMIY-TADQIQOT INSTITUTI
HUZURIDAGI ILMIY DARAJALAR BERUVCHI
DSc.13/30.12.2021.T.142.01 RAQAMLI ILMIY KENGASH**

O'ZBEKISTON MILLIY UNIVERSITETI

MAXAROV QODIRBEK TOLKUNOVICH

**BERILGANLARDA NOANIQLIK BO'LGANDA
QAROR QABUL QILISH TIZIMI**

05.01.11 – Raqamli texnologiyalar va sun'iy intellekt

**TEXNIKA FANLARI BO'YICHA FALSAFA DOKTORI (PhD)
DISSERTATSIYASI AVTOREFERATI**

Toshkent – 2024



Texnika fanlari bo'yicha falsafa doktori (PhD) dissertatsiyasi mavzusi O'zbekiston Respublikasi Vazirlar Mahkamasi huzuridagi Oliy attestatsiya komissiyasida B2023.4.PhD/T4194 raqam bilan ro'yxatga olingan.

Dissertatsiya Mirzo Ulug'bek nomidagi O'zbekiston Milliy universitetida bajarilgan.
Dissertatsiya avtoreferati uch tilda (o'zbek, rus, ingliz (rezyume)) Ilmiy kengashning veb-sahifasida (www.airi.uz) va "Ziyonet" Axborot ta'lim portalida (www.ziyonet.uz) joylashtirilgan.

Ilmiy rahbar: **Mdraximov Shavkat Fayzullayevich**
texnika fanlari doktori, professor

Rasmiy opponentlar: **Muxamediyeva Dilnoz Tulkunovna**
texnika fanlari doktori, professor

Mirzayev Nomaz
texnika fanlari doktori, professor

Yetakchi tashkilot: **Islom Karimov nomidagi**
Toshkent Davlat Texnika Universiteti

Dissertatsiya himoyasi Raqamli texnologiyalar va sun'iy intellektni rivojlantirish ilmiy-tadqiqot instituti huzuridagi DSc.13/30.12.2021.T.142.01 raqamli Ilmiy kengashning 2024 yil «06» 09 soat 15⁰⁰ dagi majlisida bo'lib o'tadi. (Manzil: 100125, Toshkent shahri, Mirzo Ulug'bek tumani, Bo'z-2 mavzesi, 17A uy. Tel.: (99871) 263-41-98, e-mail: info@airi.uz)

Dissertatsiya bilan Raqamli texnologiyalar va sun'iy intellektni rivojlantirish ilmiy-tadqiqot instituti Axborot-resurs markazida tanishish mumkin (45 raqam bilan ro'yxatga olingan). (100125, Toshkent shahri, Mirzo Ulug'bek tumani, Bo'z-2 mavzesi, 17A uy. Tel.: (99871) 263-41-98).

Dissertatsiya avtoreferati 2024 yil «23» 08 kuni tarqatildi.
(2024 yil «22» 07 dagi 11 raqamli reestr bayonnomasi)



[Handwritten signature]

[Handwritten signature]

N.Ravshanov
Ilmiy darajalar beruvchi
ilmiy kengash raisi,
texnika fanlari doktori, professor

F.M.Nuraliyev
Ilmiy darajalar beruvchi
ilmiy kengash ilmiy kotibi,
texnika fanlari doktori, professor

Sh.X.Fozilov
Ilmiy darajalar beruvchi ilmiy
kengash qoshidagi ilmiy seminar raisi,
texnika fanlari doktori, professor

KIRISH (falsafa doktori (PhD) dissertatsiyasining annotatsiyasi)

Dissertatsiya mavzusining dolzarbligi va zarurati. Jahonda kompyuter texnologiyalari jadal suratlarda rivojlanishi va ularning inson faoliyatining turli sohalarda qo'llanilishi bilan bog'liq holda berilganlarga sifatli (intellektual) ishlov berishga ehtiyoj kuzatilmoqda. Jamiyatning innovatsion rivojlanishidagi asosiy omil – bu bilimdir. Dunyoning rivojlangan mamlakatlari, jumladan AQSH, Buyuk Britaniya, Germaniya, Fransiya, Rossiya Federatsiyasi, Yaponiya, Xitoy, Janubiy Koreya, hamda Amazon, Alibaba, Google, Yandex va Microsoft kabi xalqaro kompaniyalarda prognoz qilish, regressiya va tasniflash (klassifikatsiya) masalalarini yechishga yo'naltirilgan sun'iy intellekt tizimlarini yaratish doirasida berilganlarga intellektual ishlov berish usullarini ishlab chiqish va qo'llash bo'yicha faol ilmiy tadqiqotlar olib borilmoqda. Mavjud bilimlarni o'zlashtirish bilan bir qatorda, yangi qonunlar va qonuniyatlarni ochish hisobiga ularning salmog'ini oshirib borishga katta e'tibor qaratilmoqda.

Hozirgi kunda jahonda zaif tuzilmali predmet sohalardagi bilimlarga asoslangan informatsion modellarni qurish maqsadida berilganlarga dastlabki ishlov berish usullari keng tadqiq etilmoqda. Noto'g'ri berilganlarni erta tozalash turli sohalarda berilganlarni tahlil qilishda jiddiy muammoga aylangan. Ko'plab o'tkazilgan so'rovlar tahlillari natijasiga ko'ra barcha berilganlarning 38-39 foizida sifat ko'rsatkichi past. Sifatsiz berilganlarning tahliliga ko'ra, ularning 14-15 foizi dublikatlar, 9-11 foizi noto'g'ri qiymatlar, 6-7 foizi to'liq bo'lmagan berilganlar, qolganlari noto'g'ri elektron pochta manzillari, turar joy ma'lumotlarining mos kelmasligi, formatning xatosi va h.k. Mashinali o'rgatish uchun tanlanmani shakllantirishda hal qilinmagan muammo sifatida obyektlar tavsifidagi alomatlar juftligi bo'yicha qiymatlarning mumkinligini tekshirishning mavjud emasligini keltirish mumkin. Bu muammoning o'ziga xosligi predmet soha, berilganlar to'g'riligini tekshirish uchun latent parametrlarni shakllantirish, xatolarni aniqlash uchun intellektual tizim yaratish bilan bog'liq. Shu sababli, noto'g'ri berilganlarni erta tozalash, o'tkazib yuborilgan berilganlarni qayta ishlash hamda keng doiradagi foydalanuvchilarga mo'ljallangan qulay interfeysga ega dasturlar majmuini yaratish maqsadli ilmiy tadqiqotlardan hisoblanadi.

Mamlakatimizda sun'iy intellekt texnologiyalarining katta ahamiyati va salohiyatini e'tirof etgan holda unga asoslangan tizimlarni ishlab chiqish va amaliyotga joriy etishga alohida e'tibor qaratilmoqda. Xususan, "Sun'iy intellekt uchun davlat ma'lumotlarini shakllantirishga mas'ul tegishli vazirlik va idoralarning rahbarlari o'ziga taalluqli qismlarda ma'lumotlarni o'z vaqtida joylashtirish, shuningdek, ularning to'liqligi, dolzarbligi hamda ishonchiligi yuzasidan shaxsan javobgar..."¹ kabi vazifalar belgilangan.

O'zbekiston Respublikasi Prezidentining 2017 yil 7 fevraldagi PF-4947-son "O'zbekiston Respublikasini yanada rivojlantirish bo'yicha Harakatlar strategiyasi to'g'risida"gi, 2018 yil 19 fevraldagi PF-5349-son "Axborot texnologiyalari va

¹ O'zbekiston Respublikasi Prezidentining 2021 yil 18 fevraldagi "Sun'iy intellekt texnologiyalarini jadal joriy etish uchun shart-sharoitlar yaratish chora-tadbirlari to'g'risida"gi PQ-4996-son qarori

kommunikatsiyalari sohasini yanada takomillashtirish chora-tadbirlari to'g'risida"gi va 2022 yil 28 yanvardagi PF-60-son "2022-2026 yillarga mo'ljallangan Yangi O'zbekistonning taraqqiyot strategiyasi to'g'risida"gi Farmonlari, 2018 yil 2 avgustdagi PQ-3894-son "O'zbekiston Respublikasida sog'liqni saqlashni boshqarishning innovatsion modelini joriy etish chora-tadbirlari to'g'risida"gi, 2020 yil 6 oktyabrdagi PQ-4851-son "Axborot texnologiyalari sohasida ta'lim tizimini yanada takomillashtirish, ilmiy tadqiqotlarni rivojlantirish va ularni IT-industriya bilan integratsiya qilish chora-tadbirlari to'g'risida"gi, 2021 yil 17 fevraldagi PQ-4996-son "Sun'iy intellekt texnologiyalarini jadal joriy etish uchun shart-sharoitlar yaratish chora-tadbirlari to'g'risida"gi va 2022 yil 22 avgustdagi PQ-357-son "2022-2023 yillarda axborot-kommunikatsiya texnologiyalari sohasini yangi bosqichga olib chiqish chora-tadbirlari to'g'risida"gi Qarorlari hamda axborot texnologiyalari sohasiga tegishli boshqa me'yoriy-huquqiy hujjatlarda belgilangan vazifalarni amalga oshirishga ushbu dissertatsiya tadqiqoti muayyan darajada xizmat qiladi.

Tadqiqotning respublika fan va texnologiyalari rivojlanishining ustuvor yo'nalishiga mosligi. Mazkur tadqiqot ishi O'zbekiston Respublikasi fan va texnologiyalar rivojlanishining IV. «Axborotlashtirish va axborot-kommunikatsiya texnologiyalarini rivojlantirish» ustuvor yo'nalishi doirasida bajarilgan.

Muammoning o'rganilganlik darajasi. Jahonda berilganlardagi noaniqliklarni tahlil qilish, berilganlar sifatini oshirish, berilganlar bazasidan yashirin qonuniyatlarni aniqlash, to'liq bo'lmagan berilganlarni tahlil qilish bo'yicha bir qator tadqiqotlar olib borilgan. Jumladan, Yu.I.Juravlev, V.A.Dyuk, V.N.Vapnik, K.V.Vorontsov, T.A.Gavrilova, A.Zimek, G.Litens, E.Ram, S.A.Zaxin, P.J.Garsiya-Laensina, V.Pedrits, J.Liu va boshqa xorijiy olimlarning bu boradagi ilmiy ishlari qiyosiy o'rganib chiqildi. Yu.I.Juravlev obrazlarni anglash masalasiga algebraik yondashuv nazariyasini ishlab chiqdi, V.A.Dyuk ishlarida berilganlardan mantiqiy qonuniyatlarni izlash algoritmlarining kombinatorik murakkabligini kamaytirish usullari taklif qilingan, S.A.Zaxin, P.J.Garsiya-Laensinalar to'liq bo'lmagan berilganlarda tasniflash masalalarini ishlash bo'yicha tadqiqotlar olib borgan, V.Pedrits anomal obyektlarni aniqlashda yaqin qo'shni algoritmlariga tayangan yakkalovchi o'rmon usullarini taklif qilgan, E.Ram berilganlarni tozalash, sifatli berilganlar muammolarini o'rgangan va ularni turli mezonlar bo'yicha tasniflagan.

O'zbekistonda obrazlarni anglash, intellektual tizimlarda bilimlarga ishlov berish, yashirin qonuniyatlarni aniqlash, alomatlar fazosi o'lchamini qisqartirish va olingan natijalarning samaradorligini aniqlash bo'yicha muvaffaqiyatga erishgan bir qator olimlarni ko'rsatish mumkin, jumladan, akademik M.M.Kamilov, T.F.Bekmuratov, F.T.Adilova, Sh.H.Fazilov, D.T.Muxamadiyeva, N.S.Mamatov, N.A.Ignatev, M.M.Musayev va boshqalar ushbu sohalar rivojlanishiga o'z hissalarini qo'shib kelishmoqda.

Bugungi kunda real masalalarda berilganlarda turli noaniqliklar, xususan, to'liq bo'lmagan berilganlar va alomatlarda mumkin bo'lmagan qiymatlarning uchrashi mavjud. Ushbu yo'nalish bo'yicha o'tkazilgan tadqiqotlar shuni ko'rsatadiki, berilganlardagi to'liq bo'lmagan va mantiqiy mos kelmaydigan

qiymatlari bo'lgan holda intellektual qaror qabul qilish texnologiyalari (algoritm(lari)) yetarlicha tadqiq etilmagan.

Dissertatsiya tadqiqotining dissertatsiya bajarilgan muassasaning ilmiy-tadqiqot ishlari rejalari bilan bog'liqligi. Dissertatsiya tadqiqoti Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti ilmiy-tadqiqot ishlari rejasining OT-Atex-2018-224 "Yo'g'on ichak yallig'lanishi va kolorekal saratonining rivojlanishi, kechishi va oqibatlarini prognoz qilish, hamda etiopatogenez va tavakkalchilik faktorlarini inobatga olgan holda diagnostikani amalga oshirish va davolashning axborot modelini ishlab chiqish" (2018-2021) mavzusidagi loyihasi doirasida bajarilgan.

Tadqiqotning maqsadi tanlanma obyektlari tavsifidagi mantiqiy mos kelmaydigan qiymatlarni aniqlash va to'liq bo'lmagan berilganlar mavjud holatlarda tasniflash masalasini yechishning algoritm(lari) hamda dasturiy majmuasini (tizimini) ishlab chiqishdan iborat.

Tadqiqotning vazifalari:

obyektlar tavsifidagi miqdoriy alomatlar juftliklari bo'yicha berilganlardagi mantiqiy mos kelmaslikni aniqlash algoritmini ishlab chiqish;

to'liq bo'lmagan berilganlarda alomatlar to'plamining informativligini baholash algoritmini ishlab chiqish;

to'liq bo'lmagan berilganlarni hisobga olgan holda umumlashgan baholar asosida obyektlarni tasniflash uchun alomatlar fazosini qurish;

alomatlar fazosini qisqartirish jarayonida parallel dasturlash paradigmalarini qo'llash;

ishlab chiqilgan algoritm(lar) asosida amaliy masalalarni yechishga mo'ljallangan dasturiy majmuani yaratish.

Tadqiqot obyekti sifatida berilganlarga intellektual ishlov berish usullari asosida qaror qabul qiluvchi informatsion modellar olingan.

Tadqiqot predmeti berilganlarda mantiqiy mos kelmaslikni izlash, alomatlar juftligi uchun mumkin bo'lgan qiymatlar chegaralarini hisoblash, alomatlarning turg'unligini baholash, to'liq bo'lmagan berilganlarda umumlashgan baholash usuli yordamida alomatlar fazosini qisqartirishga asoslangan algoritm(lar)dan iborat.

Tadqiqot usullari. Tadqiqotni olib borish jarayonida berilganlarni intellektual tahlil qilish, mashinali o'rgatish, obyektga yo'naltirilgan dasturlash, parallel dasturlash, sonli natijalarni vizuallashtirish usullaridan foydalanildi.

Tadqiqotning ilmiy yangiliklari quyidagilardan iborat:

miqdoriy alomatlar juftligi bo'yicha mantiqiy mos kelmaslikni aniqlash algoritmi ishlab chiqilgan;

miqdoriy alomatlar juftliklari bo'yicha mumkin bo'lgan qiymatlarni hisobga olgan holda bilimlar bazasini shakllantirish uslubiyati ishlab chiqilgan;

turg'unlik o'lchovi asosida alomatlar to'plamining informativligini baholash algoritmi va qaror qabul qilish tizimi ishlab chiqilgan;

berilganlarda noaniqlik bo'lgan holatlarni bartaraf etish uchun umumlashgan baholar asosida alomatlar fazosi o'lchamini qisqartirish algoritmi ishlab chiqilgan.

Tadqiqotning amaliy natijalari quyidagilardan iborat:

predmet soha uchun anomal obyektlar bo'yicha qaror qabul qilishni qo'llab-quvvatlash metodologiyasi taklif qilingan;

to'liq bo'lmagan berilganlarda tasniflash muammosini yangi, qisqartirilgan alomatlar fazosini shakllantirish orqali yechish taklif qilingan;

berilganlarda noaniqlik bo'lganda dastlabki ishlov berishga asoslangan qaror qabul qilishning IDEF modeli ishlab chiqilgan;

latent alomatlar fazosini shakllantirishda vaqt samaradorligiga erishish uchun parallel algoritmlar taklif qilingan.

Tadqiqot natijalarining ishonchliligi. Tadqiqot natijalarining ishonchliligi predmet sohalar tanlanmalarida hamda Internet global tarmog'ining maxsus saytlarida joylashtirilgan repozitoriylardan olingan berilganlar bazalarida o'tkazilgan hisoblash tajribalari, turg'unlik mezonidan foydalanish, tadqiqot sohasidagi mutaxassis bo'lgan boshqa mualliflarning tadqiqotlarida olingan natijalar bilan qiyosiy taqqoslash orqali izohlanadi.

Tadqiqot natijalarining ilmiy va amaliy ahamiyati. Tadqiqot natijalarining ilmiy ahamiyati tanlanma obyektlari tavsifidagi miqdoriy alomatlar juftligi uchun mumkin bo'lmagan qiymatlarni aniqlash va to'liq bo'lmagan berilganlar uchun umumlashgan baholardan foydalangan holda alomatlar fazosi o'lchamini qisqartirishga asoslangan qaror qabul qilish uslubiyati ishlab chiqilganligi bilan izohlanadi.

Tadqiqot natijalarining amaliy ahamiyati ishlab chiqilgan algoritmlar va ular asosida yaratilgan dasturiy ta'minotdan turli predmet sohalarda to'liq bo'lmagan berilganlarda tasniflash masalasini yechishda, alomatlar juftligi bo'yicha mantiqiy mos kelmaydigan qiymatlarni aniqlash maqsadida foydalanish mumkinligi hamda soha mutaxassislari uchun qaror qabul qilishda yordam beruvchi dasturiy vositalar yaratilganligi va konkret masalalarga qo'llanilganligi bilan izohlanadi.

Tadqiqot natijalarining joriy qilinishi. Berilganlarda noaniqlik bo'lganda qaror qabul qilish uchun ishlab chiqilgan algoritmlar va ularning dasturiy ta'minoti yordamida olingan ilmiy yangiliklar asosida:

miqdoriy alomatlar juftliklari uchun mumkin bo'lgan qiymatlar bo'yicha intervallarni hisoblash algoritmi asosida mantiqiy mos kelmaydigan qiymatlarni aniqlash, berilganlarda noaniqliklar bo'lgan holatlarda qabul qilinadigan qarorlarning ishonchliligini baholash algoritmlari Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti huzuridagi intellektual dasturiy tizimlar ilmiy-amaliy markazida joriy qilingan (O'zbekiston Respublikasi Raqamli texnologiyalar vazirligining 2024 yil 16 maydagi № 33-8/3190-son ma'lumotnomasi). Natijada o'pka sili bo'yicha berilganlar bazasini shakllantirish jarayonida berilganlarni tekshirish va qaror qabul qilish jarayoni vaqtini 10-15% ga qisqartirish imkonini bergan;

berilganlarda noaniqlik bo'lgan holatlarni bartaraf etish uchun umumlashgan baholar asosida alomatlar fazosini qisqartirish algoritmi va latent alomatlar fazosini shakllantirish jarayonining parallel algoritmi asosida yaratilgan qaror qabul qilish tizimi "UNICON.UZ" – Fan-texnika va marketing tadqiqotlari markazida tadqiq etilgan (O'zbekiston Respublikasi Raqamli texnologiyalar vazirligining 2024 yil 16 maydagi № 33-8/3190-son ma'lumotnomasi). Natijada ma'lumotlarni qayta

ishlash va qaror qabul qilish jarayoni vaqtini 15-20 foizga qisqartirish imkonini bergan.

Tadqiqot natijalarining aprobatsiyasi. Dissertatsiyaning asosiy nazariy va amaliy natijalari 15 ta xalqaro va 6 ta respublika ilmiy-amaliy anjumanlarida muhokamadan o'tkazilgan.

Tadqiqot natijalarining e'lon qilinganligi. Dissertatsiya mavzusi bo'yicha jami 35 ta ilmiy ish chop etilgan, shulardan O'zbekiston Respublikasi Oliy attestatsiya komissiyasining dissertatsiyalar asosiy ilmiy natijalarini chop etish tavsiya etilgan ilmiy nashrlarda 9 ta maqola, 4 tasi respublika va 5 tasi xorijiy jurnallarda nashr etilgan. O'zbekiston Respublikasi Intellektual mulk agentligining EHM uchun dasturiy mahsulot qayd etilganligi haqida 5 ta guvohnoma olingan.

Dissertatsiyaning hajmi va tuzilishi. Dissertatsiya tarkibi kirish, to'rt bob, xulosa, foydalanilgan adabiyotlar ro'yxati va ilovalardan iborat. Dissertatsiyaning hajmi 106 betni tashkil etgan.

DISSERTATSIYANING ASOSIY MAZMUNI

Dissertatsiyaning **Kirish** qismida dissertatsiya mavzusining dolzarbligi va zaruriyati asoslangan, tadqiqotning O'zbekiston Respublikasi fan va texnologiyalar taraqqiyotining ustuvor yo'nalishlariga mosligi ko'rsatilgan. Tadqiqotning maqsad va vazifalar belgilab olingan hamda tadqiqot obyekti va predmeti aniqlangan, tadqiqotning ilmiy yangiligi va amaliy natijalari bayon qilingan, olingan natijalarning ishonchliligi asoslangan va nazariy hamda amaliy ahamiyati ochib berilgan, tadqiqot natijalari amaliyotga joriy etilganligi, chop etilgan ishlar va dissertatsiya ishining tuzilishi haqida ma'lumotlar keltirib o'tilgan.

Dissertatsiyaning **“Berilganlarga dastlabki ishlov berish”** deb nomlangan birinchi bobida berilganlar va ularning xossalari, alomatlar fazosini shakllantirish, berilganlarga dastlabki ishlov berishning usul va algoritmlari qiyosiy tahlil etilgan. Xususan, berilganlarni saqlash vositalari va ko'rinishlari, alomatlarining mumkin bo'lgan qiymatlari shkalalari keltirilgan. Ayni paytda ma'lumotlar deyarli hamma joyda yaratiladi: seysmik sensorlarning mintaqaviy tarmoqlari, tibbiy muassasalarda bemorlarning holati to'g'risidagi so'rovnomalar va yozuvlar, muayyan mavzu bo'yicha jamoatchilik fikrini o'rganish, mobil aloqa provayderlaridagi mijozlar faoliyati to'g'risidagi ma'lumotlar bazasi va boshqalar. Ushbu yig'ilayotgan real dunyo ma'lumotlarining aksariyati umumiy kamchilikka ega - yetishmayotgan yoki noaniq ma'lumotlar. Masalan, sanoat tajribasida ma'lumotlarni yig'ish jarayonida mexanik/elektron nosozliklar tufayli ba'zi natijalar bo'lmasligi mumkin, tibbiy diagnostikada ba'zi tekshiruvlarni o'tkazib bo'lmaydi, chunki shifoxonada kerakli tibbiy asbob-uskunalar mavjud emas yoki ayrim tibbiy tekshiruvlar ma'lum bemorlarga mos kelmasligi mumkin. Yana bir misol, har xil turdagi testlarni o'tkazadigan shifokor tomonidan tekshiruv bo'lishi mumkin; ba'zi testlarning natijalari bir zumda mavjud bo'lishi, boshqalari esa bir necha kun talab qilishi mumkin. Har qanday holatda, faqat mavjud natijalardan foydalangan holda, darhol dastlabki tashxis qo'yish kerak bo'lishi mumkin. Bunda uch xil yondashuvdan foydalaniladi: noaniq qiymatlari bo'lgan obyektlar yoki

alomatlarni o‘chirib tashlash; noaniq qiymatlarni hisoblash orqali to‘ldirish; noaniq qiymatlar bilan masalani yechish. Birinchi ikkita yondashuv bir qancha kamchiliklarga ega, shu sababli noaniq qiymatlar bilan masalani yechishning usul va algoritmlarini ishlab chiqish dolzarb vazifalardan biri.

Dissertatsiyaning “**Alomatlarning mumkin bo‘lgan qiymatlarini nazorat qilish algoritmi**” deb nomlangan ikkinchi bobida predmet sohada alomatlar juftliklari bo‘yicha mumkin bo‘lmagan qiymatlarni aniqlash usuli taklif qilingan va undan foydalanib kiruvchi qiymatlarda xatolarni aniqlashga mo‘ljallangan algoritmlar taklif qilingan.

$X = (x_1, \dots, x_n)$ miqdoriy alomatlar bilan tavsiflanuvchi $E_0 = \{S_1, \dots, S_m\}$ obyektlar to‘plami $A = \{a_{uv}\}_{m \times n}$ jadval ko‘rinishida berilgan bo‘lsin. Har bir $(x_i, x_j) \in X(n)$ alomatlar juftligi bo‘yicha quyidagilarni aniqlash talab qilinadi:

- mumkin bo‘lgan $[z_1, z_2]$ interval chegaralarini quyidagi formula asosida:

$$R(i, j) = \left\{ \frac{a_{ki}}{P_i} - \frac{a_{kj}}{P_j} \right\}_{k \in \{1, \dots, m\}}, \quad (1)$$

bu yerda $S_k = (a_{k1}, \dots, a_{kn})$, $S_k \in E_0$, P_i, P_j – mos ravishda E_0 to‘plamdagi x_i va x_j alomatlar medianalari qiymatlari;

- z_1 va z_2 – interval chegaralari qiymatlari:

$$z_1 = \min_{E_0} R(i, j), z_2 = \max_{E_0} R(i, j). \quad (2)$$

Shuni ta’kidlash kerakki, faqatgina soha mutaxassisi alomatlar juftligi qiymatlari mantiqiy to‘g‘riligini baholay oladi. Ushbu usulni dastlabki ishlov berishda mos ekspertni jalb qilish orqali quyidagicha amalga oshirish mumkin:

1. Har bir (x_i, x_j) , $i < j$, $1 \leq i, j \leq n$ alomatlar juftligi uchun $Z = \{[z_1^{ij}, z_2^{ij}]\}$ intervallar to‘plamini qurish;
2. Interval chegarasida yotuvchi obyektlar to‘plamini qurish:

$$\Pi = \bigcup_{i, j \in I, i \neq j} \left\{ S_u, S_v \left| \frac{a_{ui}}{P_i} - \frac{a_{uj}}{P_j} = z_1 \text{ ba } \frac{a_{vi}}{P_i} - \frac{a_{vj}}{P_j} = z_2 \right. \right\};$$

3. $\forall S_u \in \Pi$ obyekt uchun qiymatlar mumkinligining ekspert xulosasini qabul qilish (o‘zgarishsiz qoldirish, berilgan qiymatlarni to‘g‘rilash yoki obyektни o‘chirish);
4. Agar tanlanma o‘zgartirilgan bo‘lsa (1) va (2) ga ko‘ra mumkin bo‘lgan qiymatlar uchun interval chegaralarini qayta hisoblash.

Algoritm tomonidan shakllantirilgan Z intervallar bazasidan qo‘yilgan masalani yechishda yangi obyektни tavsiflashdagi qiymatlarning mantiqiy mos kelishini tekshirishda foydalanish mumkin. Bunda mumkin bo‘lmagan obyektни aniqlashda ikki turdagi xato keltirilishi mumkin:

- “*kritik xato*” – alomatlar juftligi bo‘yicha qiymati interval chegarasida yoki chegarasidan tashqarida yotgan obyekt uchun;

- “*ogohlantirish*” – alomatlar juftligi bo‘yicha qiymati interval chegarasiga “*xavfli*” darajada yaqin joylashgan obyekt uchun.

Hisoblash eksperimenti uchun Nancy Howell tomonidan to'plangan "Kalahari !Kung San people" tanlanmasi olindi. Ushbu tanlanmadan faqatgina odamning yoshi, bo'y uzunligi va vazn alomatlari bo'yicha olingan juftliklarga tahlil qilindi. Tahlil natijalari 1-jadvalda keltirilgan.

1-jadval.

Alomatlar juftligi bo'yicha intervallar qiymatlari

	Alomatlar juftligi	$[z_1, z_2]$ interval	Interval chegarasida yotuvchi obyektlar (alomatlar juftligi qiymatlari)	
			chap	o'ng
1	(Bo'y, Vazn)	[-0.58, 0.46]	5 (163.83, 62.99)	359 (109.22, 11.71)
2	(Bo'y, Yosh)	[-1.90, 0.72]	164 (140.97, 85.60)	359 (109.22, 2.00)
3	(Vazn, Yosh)	[-1.79, 0.99]	222 (35.81, 82.00)	182 (61.80, 22.00)

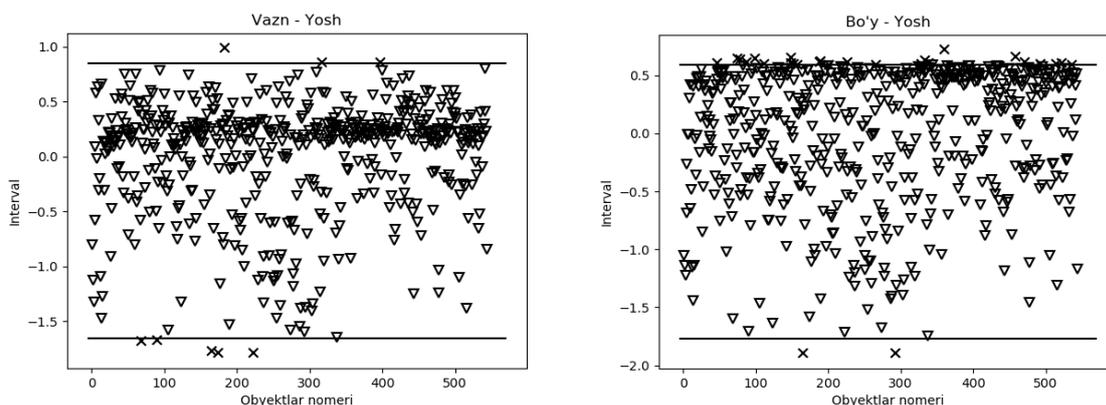
2-jadvalda mumkin bo'lmagan obyektlarni taqdim etish uchun namunalar keltirilgan: (2) bo'yicha qiymat interval chegarasidan tashqarida yoki chegaraning o'zida (kritik xato); (2) bo'yicha qiymat interval chegarasiga xavfli yaqinlikda joylashgan (ogohlantirish).

2-jadval.

Shubhali obyektlarni taqdim etish

Obyekt nomeri	Alomatlar nomlari	Alomatlar qiymatlari	Interval	Alomatlar juftligi uchun (2) qiymati	Xato statusi
6	Bo'y	163.83	[-0.41, 0.44]	-0.47	<i>Kritik xato</i>
	Vazn	62.99			
90	Bo'y	136.53	[-2.23, 0.66]	-2.01	<i>Ogohlantirish</i>
	Yosh	79.00			
90	Vazn	36.29	[-2.15, 0.73]	-2.02	<i>Ogohlantirish</i>
	Yosh	79.00			
165	Vazn	40.94	[-2.14, 0.73]	-2.15	<i>Kritik xato</i>
	Yosh	85.60			
174	Vazn	32.21	[-2.15, 0.73]	-2.12	<i>Ogohlantirish</i>
	Yosh	79.00			

1-rasmda taklif qilingan usul bo'yicha alomatlar juftligi bo'yicha qurilgan interval orasida obyektlarning joylashuvi ko'rsatilgan. "X" belgisi bilan alomatlar juftligi bo'yicha mantiqiy mos bo'lmagan qiymatlari bo'lishi mumkin bo'lgan obyektlar (anomal obyektlar) ajratib ko'rsatilgan. Bunda anomal obyektlar sifatida interval chegarasidan $5h$ masofada yotgan obyektlar qaralgan, $h = (z_2 - z_1) / 100$.



1-rasm. Alomatlar juftligi bo'yicha obyektlarning grafik ko'rinishi

Anomal obyektlarni aniqlash usullarini taqqoslash uchun boshlang'ich (asl) berilganlarga sun'iy ravishda 20 ta anomal obyekt qo'shib hisoblash eksperimenti o'tkazildi. Taklif etilayotgan usul alomatlar juftligi bilan ishlaganligi sababli eksperiment jarayonida obyekt bir juft alomatlarga nisbatan bo'lsa ham anomal obyekt deb hisoblandi. Taqqoslashni yanada aniqroq taqdim etish maqsadida aniqlik (precision) va f_1 -bahodan (f_1 -score) foydalanildi. Taqqoslash natijalari 3-jadvalda keltirilgan.

3-jadval.

Anomal obyektlarni aniqlash bo'yicha solishtirma natijalar

Test №	Isolation Forest		DBSCAN		Taklif etilayotgan usul	
	F_1 baho	Aniqlik	F_1 baho	Aniqlik	F_1 baho	Aniqlik
Test 1	0.70	0.70	0.92	0.90	0.85	1.00
Test 2	0.80	0.80	0.90	0.86	0.91	1.00
Test 3	0.75	0.75	0.90	0.86	0.88	1.00
Test 4	0.70	0.70	0.90	0.86	0.91	1.00
Test 5	0.70	0.70	0.92	0.92	0.72	0.62
Test 6	0.65	0.65	0.90	0.90	0.88	1.00
Test 7	0.60	0.60	0.90	0.90	0.91	1.00
Test 8	0.70	0.70	0.88	0.82	0.82	0.77
Test 9	0.75	0.75	0.88	0.82	0.91	1.00
Test 10	0.70	0.70	0.85	0.81	0.88	1.00
O'rtacha	0.70	0.70	0.89	0.86	0.87	0.94

3-jadvalga ko'ra, DBSCAN usuli natijalarida, ko'plab testlarda f_1 -bahosi tavsiya etilgan usuldan oshib ketgan, ammo tavsiya etilgan usulda aniqlik yuqori natija bergan. Aslida, qo'yilgan vazifa anomal obyektlarni topish bo'lgani uchun ham aniqlik qiymati (to'g'ri topilgan (haqiqiy) ijobiy qiymatlar nisbati) bo'yicha taklif etilayotgan usul orqali ko'proq anomal obyektlarni topish mumkin degan xulosaga kelish mumkin.

Dissertatsiyaning **“To'liq bo'lmagan berilganlarda qaror qabul qilish va uning ishonchligini baholash algoritmlari”** nomli uchinchi bobida to'liqmas berilganlarda alomatlarining turg'unlik xossasi, alomatlar to'plamining informativligini baholash usuli, umumlashgan baxo asosida alomatlar fazosini

qisqartirish usuli va tasniflash algoritmlarini qo'llash hamda sinfga ajratish aniqligini baholash natijalari keltirilgan.

Masalaning qo'yilishi. Obyektlari o'zaro kesishmaydigan K_1 va K_2 sinflarga bo'lingan $E_0 = \{S_1, \dots, S_m\}$ to'plam (tanlanma) berilgan bo'lsin. Tanlanma obyektlari n ta turli toifadagi alomatlar to'plami $X(n) = \{x_1, \dots, x_n\}$ bilan tavsiflangan bo'lib, ularning δ ($\delta < n$) tasi nominal, $n - \delta$ tasi miqdoriy shkalada o'lchanadi. Berilganlarda o'tkazib yuborilgan qiymatlar mavjud deb qaraladi. O'ng'aylik uchun, K_1 sinf vakillarini ro'y bergan holatlar (holatlar) va K_2 – ro'y bermagan holatlar (no holatlar) deb hisoblanadi. Ikki sinfli masala qaralishiga sabablardan biri – har qanday obyektning umumlashgan bahosi nisbiydir, u qarama-qarshi sinf obyektlariga qiyoslash natijasida yuzaga keladi. Ikkinchidan, har qanday k ($k > 2$) sinfli masalani ikki sinfli masalalar kaskadi ko'rinishida yechish mumkin. $X(n)$ ga tegishli alomatlarning o'lchangan qiymatlari to'plamida aniqlangan manfiy bo'lmagan o'lchov – $\mu(x_i)$ aniqlangan bo'lsin. Talab qilinadi:

- miqdoriy alomatlar uchun ustunlik intervallarini hisoblash;
- intervallarning (nominal alomatlar uchun har bir gradatsiyaning) K_1 sinfga tegishlilik funksiyasi qiymatlarini hisoblash;
- o'lchangan qiymatlar soniga bog'liq ravishda $\mu(x_i)$ qiymatlaridagi tarqalishlarni (farqlanishlarni) baholash;
- $\mu(x_i)$ qiymatlari asosida alomatlar to'plamining informativligini hisoblash;
- tegishlilik funksiyasi asosida alomatlar fazosini noxiziq o'girish;
- ko'rsatilgan alomatlar to'plamlari asosida obyektlarning umumlashgan baholarini hisoblash orqali yangi alomatlar fazosini qurish;
- yangi alomatlar fazosida tasniflash aniqligini baholash.

Turg'unlik qiymatini baholashning obyektivligini isbotlash uchun obyektlar tavsifidagi bo'shliqlarni tasodifiy taqsimlash bilan tajriba o'tkazildi. Har bir alomat bo'yicha o'tkazib yuborilgan qiymatlar foizi 0.0% dan 35.0% gacha bo'lgan holatlar tahlil qilindi. O'tkazib yuborilgan qiymatlarning alomatning turg'unligiga ta'sirini tadqiq qilish bo'yicha eksperiment natijalarining bir qismi 4-jadvalda keltirilgan. Jadvaldagi dispersiya qiymatlarining tahlili o'tkazib yuborilgan qiymatlar bo'lganda ham turg'unlik qiymatining past o'zgaruvchanligi haqida xulosa qilish mumkin.

4-jadval.

Turg'unlik qiymatlari bo'yicha eksperiment natijalari

Alomat nomi	Matematik kutilma	Dispersiya	Alomat nomi	Matematik kutilma	Dispersiya
Region	0.6016	0.000068	Normal Chair	0.5339	0.000250
Occupation	0.6094	0.000105	Constipation	0.5566	0.000064
Age	0.6204	0.000174	Stool Frequency	0.5351	0.000176
Gender	0.5175	0.000244	Chair Character	0.5380	0.000100
Pyrazinamide	0.5464	0.000052	Eosinophils	0.5862	0.000124
Ethambutol	0.5111	0.000074	Lymphocytes	0.6015	0.000760
...

Alomatlarning matematik kutilmasi qiymatlarini o'smaydigan tartibdagi ketma-ketligini

$$\omega_1(t_1), \dots, \omega_g(t_g), \dots, \omega_n(t_n), \quad t_i \in \{1, \dots, n\} \quad (3)$$

orqali belgilaylik. (3) ketma-ketlik elementlarining indekslarini alomat rangi deb qaraladi. Ixtiyoriy mumkin bo'lgan S obyekt tavsifidagi qiymati mavjud bo'lgan k ($[n/2] \leq k < n$) ta alomatlar to'plam ostisining informativligini aniqlashni ko'raylik. 4-jadvaldagi qiymatlar asosida alomatlarning tartiblangan ketma-ketligi quyidagi ko'rinishda bo'ladi:

$$\begin{aligned} & x_{48}, x_{14}, x_{16}, x_{15}, x_{40}, x_{36}, x_{38}, x_1, x_3, x_{32}, x_{39}, x_2, x_{47}, x_{43}, x_{45}, x_{44}, \\ & x_{46}, x_{18}, x_5, x_{42}, x_{17}, x_9, x_{26}, x_{20}, x_{23}, x_{25}, x_{24}, x_8, x_{33}, x_{34}, x_{27}, x_{19}, \\ & x_{22}, x_{28}, x_{29}, x_{35}, x_{41}, x_{12}, x_{37}, x_6, x_{31}, x_{21}, x_{13}, x_{30}, x_{11}, x_4, x_{10}, x_7. \end{aligned} \quad (4)$$

Aytaylik, $D(S, k)$ – S obyektning o'lgangan qiymatli k ta alomatlarining ranglarining yig'indisi bo'lsin. Yig'indi qiymati $k(1+k)/2 \leq D(S, k) \leq k(2n-k+1)/2$ oralig'ida bo'ladi. S obyekt tavsifidagi qiymati mavjud bo'lgan k ta ($[n/2] \leq k < n$) alomatlar to'plam ostisining informativligini baholash uchun

$$\Omega(S, k) = 1 - \left(\frac{D(S, k) - \alpha}{\beta - \alpha} \right) \left(1 - \frac{k}{n} \right) \quad (5)$$

o'lchovidan foydalanish taklif qilinadi. Bu yerda $\alpha = k(1+k)/2$, $\beta = k(2n-k+1)/2$. (5) o'lchovni ixtiyoriy obyekt tavsifidagi berilganlarni tahlil qilishda qo'llash tavsiya qilinadi. Masalan, o'tkazib yuborilgan qiymatlari mavjud bo'lgan obyekt bo'yicha qaror qabul qilishda qo'shimcha indikator (qarorga ishonch o'lchovi) sifatida ishlatilishi mumkin. 5-jadvalda (4) ga muvofiq, alomatlarning ba'zi to'plamlariga nisbatan informativlik o'lchovini qo'llash namunalari keltirilgan. Xulosa uchun informativlik o'lchovining mumkin bo'lgan qiymatlari diapazoni mutaxassis tomonidan mustaqil ravishda belgilanishi mumkin. Jadvaldagi namunada diapazonlar quyidagicha belgilangan:

- $[0, 0.6)$ – “yetarli emas”;
- $[0.6, 0.72)$ – “qoniqarli”;
- $[0.72, 0.86)$ – “yetarli”;
- $[0.86, 1.0]$ – “mos keladi”.

5-jadval.

Alomatlar to'plami uchun informativlik o'lchovi

№	Ishlatilgan alomat nomerlari	Xulosa	(5) qiymati
1	4, 6, 8, 9, 10, 11, 13, 17, 19, 21, 23, 24, 25, 27, 29, 30, 31, 33, 35, 41, 42	yetarli emas	0.5179
2	4, 6, 7, 8, 10, 11, 13, 19, 21, 22, 27, 28, 29, 30, 31, 33, 34, 35, 37, 41	yetarli emas	0.4271
3	8, 9, 12, 17, 19, 20, 22, 23, 26, 27, 29, 33, 34, 35, 36, 37, 39, 42, 44, 46	qoniqarli	0.6813
4	2, 5, 6, 8, 13, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 35, 37, 39, 41, 45	qoniqarli	0.6313

№	Ishlatilgan alomat nomerlari	Xulosa	(5) qiymati
5	1, 5, 8, 9, 17, 19, 23, 25, 26, 27, 28, 32, 33, 35, 37, 38, 40, 42, 43, 44, 46	yetarli	0.7708
6	2, 3, 5, 9, 12, 17, 18, 19, 20, 22, 25, 26, 28, 33, 35, 36, 38, 40, 43, 44, 46, 47	yetarli	0.8097
7	1, 2, 3, 5, 9, 14, 15, 17, 18, 19, 22, 23, 24, 26, 32, 33, 34, 38, 39, 42, 46, 48	mos keladi	0.8797
8	3,5, 9, 14, 15, 16, 17, 20, 26, 32, 34, 38, 39, 40, 42, 43, 44, 46, 47, 48	mos keladi	0.9365

Alomatlar qiymatlarini noxiziqli o'g'irishning mohiyati boshlang'ich qiymatlarni sinflarga tegishlilik funksiyalari qiymatlari bilan almashtirish hisoblanadi. $x_c \in X(n)$ uchun sinf obyektlari orasidagi masofa quyidagicha aniqlanadi:

$$G_c = \frac{q_1 + q_2}{2}, \quad (6)$$

bu yerda $q_1 = \min\{f_c(\gamma) \mid 1 - f_c(\gamma) < 0.5, \gamma = 1, \dots, \tau_c\}$

$q_2 = \max\{f_c(\gamma) \mid 0.5 - f_c(\gamma) > 0, \gamma = 1, \dots, \tau_c\}$.

$S_i = \{x_{iu}\}_{u \in D}$ obyekt uchun (6) dan foydalanib α_{ic} gradatsiya qiymatini hisoblashda o'lchov shkalalarini hisobga olgan holda quyidagi ikki shartning biri ko'riladi: $x_{ic} \in [r_u, r_v]^\gamma$ yoki $x_{ic} = \gamma$. α_{ic} qiymatini hisoblashda tegishlilik funksiyasidan foydalaniladi:

$$a_{ic} = \begin{cases} 1, & f_c(\gamma) < G_c, \\ 2, & f_c(\gamma) > G_c. \end{cases}$$

Obyektning umumlashgan bahosini hisoblash uchun $\{1,2\}$ gradatsiyalar bilan tavsiflanuvchi alomatlar hissasidan foydalaniladi. x_c alomatning $j \in \{1,2\}$ gradatsiya bo'yicha xissasi quyidagicha aniqlanadi:

$$\eta_c(j) = w_c \left(\frac{\alpha_{cj}^1}{T_{1c}} - \frac{\alpha_{cj}^2}{T_{2c}} \right), \quad (7)$$

bu yerda $\alpha_{cj}^1, \alpha_{cj}^2$ – x_c alomatning j gradatsiyasining mos ravishda K_1 va K_2 sinflardagi miqdori, w_c – x_c alomatning sinf ichidagi o'xshashlik va sinflararo farqlanish ko'paytmasi orqali hisoblangan vazni. Nominal o'lchov shkalasida tavsiflangan $S_r \in E_0$ obyektning umumlashgan bahosi quyidagicha hisoblanadi:

$$R(S_r) = \sum_{i \in D} \eta_c(\alpha_{ri}). \quad (8)$$

Yangi alomatlar fazosini qurishda asosiy kriteriya sifatida turg'unlik qiymati olinadi. Alomat turg'unligining qiymatlari $(0.5;1]$ intervalga tushadi. Agar turg'unlik qiymati birga teng bo'lsa, tanlanmadagi obyektlar K_1 va K_2 sinflarga xatosiz ajratiladi. Nisbatan kichik turg'unlik qiymatiga ega bo'lgan alomatlarni

olib tashlash berilganlarni tahlil qilish algoritmlarida hisoblash murakkabligini kamaytirish uchun taklif qilinadi.

Ω , $\Omega \subset X(n)$ orqali nisbatan kichik turg'unlik qiymatiga ega bo'lgan nominal alomatlar to'plami belgilansin, u holda $\Pi = \{i | i = X(n) \setminus \Omega\}$ guruhlashda ishlatiladigan alomatlar nomerlari. Latent alomatlar hosil qilishda Π dagi turli alomatlar guruhlaridan foydalaniladi. Alomatlar turg'unlik qiymati bo'yicha tartiblanadi va k tadan ($k = 2, \dots, |\Pi|/2$) guruhlanadi. k ning optimal qiymati hisoblash eksperimenti orqali aniqlanadi. Har bir alomatlar guruhlar bo'yicha (7) va (8) orqali umumlashgan baholar hisoblanadi. Umumlashgan baholar qiymatidan to'liq bo'lgan yangi alomatlar fazosida obyekt tavsifidagi latent alomatlar qiymatlari sifatida foydalaniladi.

Birinchi hisoblash eksperimenti uchun ikkita tanlanma olingan. Ulardan biri 167 obyekt, 48 ta alomat (16 tasi miqdoriy, 32 tasi nominal) bilan tavsiflangan tuberkulyoz kasalligi bemorlari bazasi. Ikkincha baza – Viskonsinda ko'krak bezi saratoni bilan kasallangan bemorlar, 569 ta obyekt, 30 ta alomat bilan tavsiflangan (barchasi miqdoriy). Har bir alomatlar uchun 0.0% dan 50.0% gacha bo'sh qiymatlar sun'iy ravishda hosil qilingan. Alomatlar turg'unlik qiymati bo'yicha tartiblanadi va k tadan guruhlanadi. k ning optimal qiymati hisoblash eksperimenti orqali aniqlandi. Yangi alomatlar fazosida tasniflash aniqligi quyidagi mashinali o'rgatish usullari orqali tekshirildi:

- Bayes klassifikatori (Naïve Bayes);
- k-yaqin qo'shnilar uchuli (k-nearest neighbors, KNN);
- tayanch vektorlar usuli (Support Vector Machine, SVM);
- tasodifiy o'rmon (Random Forest);
- qarorlar daraxti (Decision Tree);
- ko'p qatlamli perseptron (Multi-layer Perceptron, MLP).

6-jadvalda tanlanmalarga sun'iy ravishda har bir alomatlar uchun 0.0% dan 50.0% gacha bo'sh qiymatlar hosil qilingandagi tasniflash aniqligi natijalari keltirilgan ($k=5$ bo'lganda). Tasniflash aniqligi natijalarini aniqlash uchun o'nta eksperiment o'tkazilgan va ularning o'rtacha qiymati olingan. Natijalar shuni ko'rsatadiki, yangi alomatlar fazosida bo'sh qiymatlar mavjud bo'lganda ham (50% gacha) tasniflash aniqligi bo'yicha yuqori natijalarga erishilgan.

6-jadval.

Tasniflash aniqligi bo'yicha o'tkazilgan eksperiment natijalari

№	Bo'sh qiymatlar (%)	Tasniflash aniqligi					
		Naïve Bayes	KNN	SVM	Random Forest	Decision Tree	MLP
Tuberkulez kasalligi							
1	0	0.960	0.960	0.940	0.980	0.920	0.980
2	5	0.958	0.958	0.952	0.952	0.914	0.958
3	10	0.956	0.956	0.936	0.956	0.912	0.948
4	15	0.936	0.906	0.932	0.932	0.884	0.930
5	20	0.926	0.948	0.910	0.934	0.890	0.926
6	25	0.946	0.926	0.924	0.934	0.890	0.932

№	Bo'sh qiymatlar (%)	Tasniflash aniqligi					
		Naïve Bayes	KNN	SVM	Random Forest	Decision Tree	MLP
7	30	0.916	0.930	0.936	0.942	0.894	0.920
8	35	0.940	0.940	0.920	0.942	0.888	0.924
9	40	0.946	0.944	0.922	0.948	0.902	0.914
10	45	0.924	0.940	0.914	0.938	0.880	0.906
11	50	0.928	0.922	0.916	0.930	0.890	0.912
Ko'krak saratoni kasalligi							
1	0	0.976	0.977	0.977	0.977	0.942	0.959
2	5	0.970	0.966	0.971	0.968	0.945	0.958
3	10	0.966	0.967	0.971	0.965	0.952	0.956
4	15	0.969	0.959	0.965	0.967	0.946	0.958
5	20	0.968	0.961	0.965	0.964	0.934	0.954
6	25	0.966	0.964	0.969	0.963	0.950	0.946
7	30	0.956	0.960	0.955	0.961	0.940	0.944
8	35	0.963	0.954	0.958	0.957	0.923	0.950
9	40	0.960	0.954	0.963	0.960	0.929	0.949
10	45	0.964	0.957	0.967	0.960	0.936	0.948
11	50	0.958	0.953	0.964	0.950	0.931	0.949

Ikkinchi hisoblash eksperimentida taklif qilinayotgan usul boshqa yondashuvlar bilan taqqoslandi. Buning uchun "Pima hindulari diabet kasalligi" tanlanmasidan foydalanildi. Tanlanma 768 ta obyekt va 8 ta alomat bilan tavsiflangan. Obyektlarning 500 (65.1%) tasi sog'lom va 268 (34.9%) tasi kasal sinfga tegishli. Tanlanmadagi beshta alomat (glyukoza, qon bosimi, terining qalinligi, insulin va tana-vazn indeksi) qiymatlarida nol qiymati mavjud, lekin haqiqatda bunday qiymat bo'lishi mumkin emas. Shu sababli, bu qiymatlarni xato qiymatlar (o'tkazib yuborilgan qiymatlar) deb hisoblash mumkin. Natijada o'tkazib yuborilgan qiymatlar soni alomatlar bo'yicha quyidagicha taqsimlandi:

- glyukoza: 5 ta (0.65 %);
- qon bosimi: 35 ta (4.56 %);
- terining qalinligi: 227 ta (29.56 %);
- insulin: 374 ta (48.70 %);
- tana-vazn indeksi: 11 ta (1.43 %).

Umumiy hisobda esa, 376 ta obyektida o'tkazib yuborilgan qiymatlar mavjud. Bu butun tanlanmaning 48.96 foizini tashkil qiladi. 7-jadvalda mazkur tadqiqot ishi doirasida taklif etilayotgan o'tkazib yuborilgan qiymatlardan qutulish uchun yondashuv boshqa olingan natijalar va yondashuvlar bilan taqqoslangan. Taklif etilayotgan fazoni qisqartirish usulini qo'llab tasniflash algoritmlarini ishlatish aniqligi ko'p holatlarda yaxshi natija bergan. Eng yaxshi natijaga alomatlarini guruhlash parametri – $k=3$ bo'lganda va tasniflash uchun Naïve Bayes algoritmidan foydalanganda erishildi.

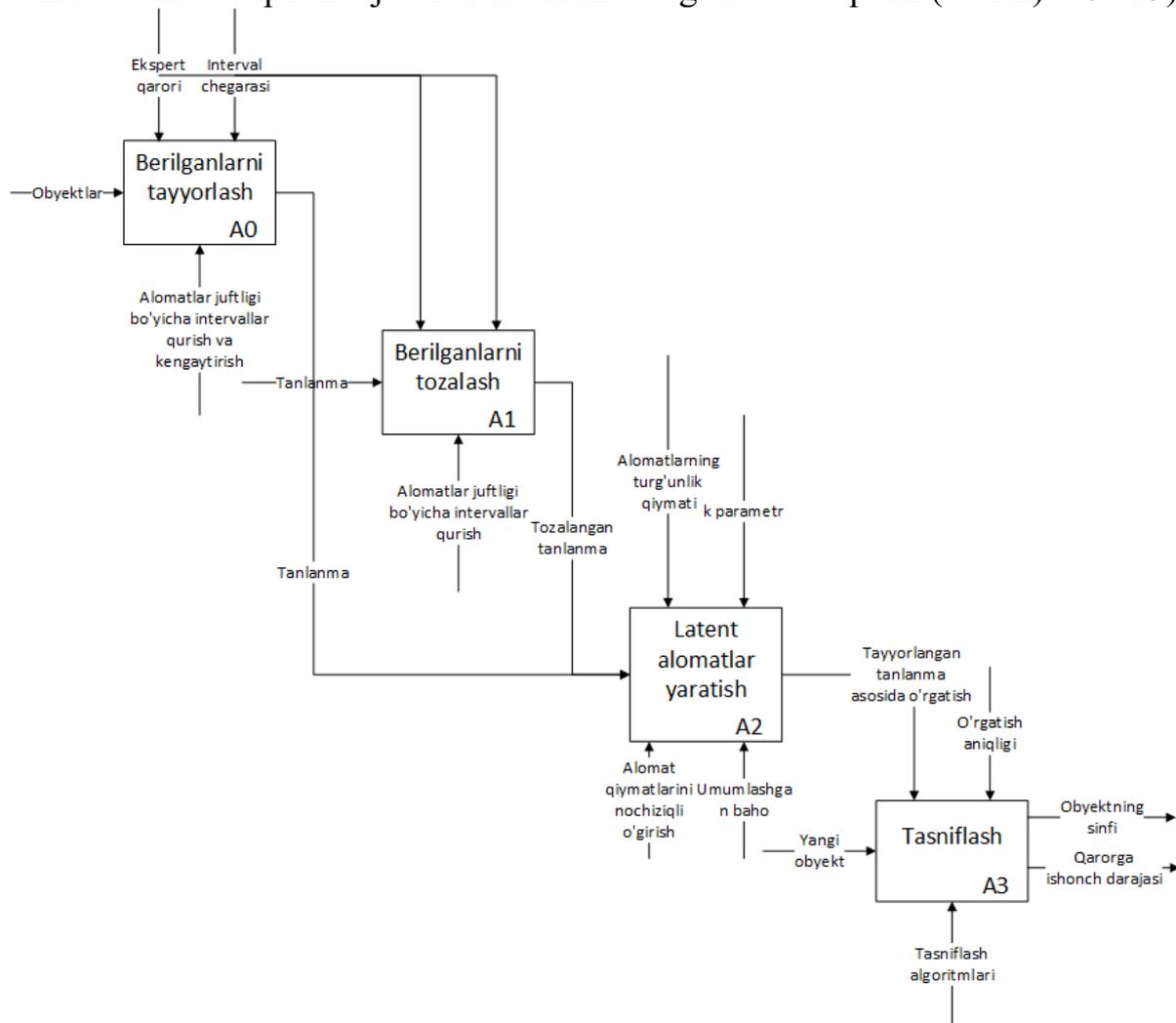
To'liq bo'lmagan berilganlarda o'tkazilgan eksperiment natijalari

№	Qo'llanilgan yondashuv	Tasniflash aniqligi					
		Naïve Bayes	KNN	SVM	Random Forest	Decision Tree	MLP
1.	To'liq bo'lmagan obyektlarni o'chirish	0.738	0.720	0.729	0.771	0.695	0.754
2.	O'rtacha qiymat bilan to'ldirish	0.740	0.709	0.745	0.745	0.692	0.745
3.	k yaqin qo'shni usuli bilan to'ldirish (k=11)	0.736	0.701	0.727	0.757	0.732	0.746
4.	Umumiy regressiya neyron tarmog'i	0.802					
5.	Gradient Boosting usuli	0.789					
6.	Logdisc usuli	0.777					
7.	IncNet usuli	0.776					
8.	Taklif qilinayotgan usul: k=2	0.761	0.799	0.785	0.777	0.757	0.788
	k=3	0.814	0.758	0.788	0.766	0.746	0.775
	k=4	0.809	0.788	0.771	0.771	0.766	0.796

Dissertatsiyaning “**Dasturiy ta'minot arxitekturasi va amaliy masalalarni yechishda qo'llanilishi**” nomli to'rtinchi bobida zamonaviy intellektual axborot tizimlari va qaror qabul qilishga ko'maklashuvchi intellektual tizimlarni yaratishda qo'llaniladigan dasturiy ta'minot arxitekturalari tadqiq qilingan va mikroservisli gibrid arxitekturadan foydalanish taklif qilingan, yaratilgan dasturiy ta'minot qanday mikroservislarni o'z ichiga olishi kerakligi sanab o'tilgan. Yaratiladigan dasturiy ta'minotning funksional modellari taklif qilingan, dasturiy ta'minotda parallellik tamoyillarini ishlatishning samaradorligi hisoblash eksperimenti orqali ko'rsatilgan, yaratilgan dasturiy ta'minotning interfeysi va amaliy masalada qo'llash namunasi keltirib o'tilgan. Dissertatsiya ishida taklif qilingan algoritmlar bo'yicha yaratiladigan dasturiy ta'minotning IDEF0 funksional (2-rasm) va IDEF1.x axborot modellari yaratilgan va funksional modellar asosida tibbiyot sohasi uchun yaratilishi mumkin bo'lgan tizimning amal qilish sxemasi keltirib o'tilgan.

Qaralayotgan ushbu tadqiqot ishida yangi (latent) alomatlar fazosini qurishda miqdoriy alomatlar uchun ustunlik intervallarini aniqlash va intervallarning sinfga tegishlilik funksiyasi qiymatini hisoblash bosqichlari to'la tanlash bilan bog'liq hisoblash murakkabligiga ega bo'lib, ularni amalga oshirishda Multiple Instruction Multiple Data arxitekturasidan foydalanish yuqori samaradorlikka olib keladi. Algoritmlarni ishlatishda parallellizm tamoyillarini qo'llash samaradorligini baholash uchun hisoblash eksperimenti o'tkazilgan. Hisoblash eksperimenti uchun

prostata raki kasalligini tashhis qilish bilan bog‘liq tanlanma olindi. Ushbu tanlanma 102 ta obyekt va 12625 ta alomatdan tashkil topgan. Eksperiment Intel(R) Core(TM) i7-8550U, 1.80GHz protsessorli kompyuterda amalga oshirildi (operativ xotira hajmi – 8Gb, grafik xotira hajmi – 2Gb). Ushbu protsessorida 4 ta jismoniy yadro va 8 ta mantiqiy yadro mavjud. Alomatlar fazosini nochiziqli o‘g‘irish jarayonini amalga oshirishda parallel hisoblashdan foydalanish natijalari 8-jadvalda keltirilgan. Eksperimentda tanlanma hajmini ko‘paytirish maqsadida alomatlar soni sun‘iy ravishda ko‘paytirilgan. Parallelizmdan foydalanishning samaradorligi Amdal qonuniga asosan keltirib o‘tilgan (9-jadval, faqatgina ketma-ket hisoblashlar orqali natija olish mumkin bo‘lgan dastur qismi (ulushi) = 0.125).



2-rasm. Tizimning IDEF0 funksional modeli

8-jadval.

Parallel hisoblash natijalari

Alomatlar soni	Protsessor yadrolari soni							
	1	2	3	4	5	6	7	8
12625	17.02	10.58	9.14	8.10	8.32	8.78	9.48	12.17
126250	158.71	97.56	73.18	65.19	61.14	65.79	63.79	68.01
252500	335.05	184.49	143.94	126.38	135.99	129.01	117.66	122.93
378750	494.55	291.08	227.54	198.09	193.54	188.15	182.27	183.58
505000	689.10	385.37	298.90	268.16	265.99	288.04	292.64	278.94

Parallel hisoblash samaradorligi

Alomatlar soni / Yadrolar soni	Tezlashish			Samaradorlik		
	2	3	4	2	3	4
Ideal holat	1.78	2.40	2.91	0.89	0.80	0.73
12625	1.61	1.86	2.10	0.80	0.62	0.53
126250	1.63	2.17	2.43	0.81	0.72	0.61
252500	1.70	2.18	2.57	0.85	0.73	0.64
378750	1.70	2.23	2.60	0.85	0.74	0.65
505000	1.74	2.31	2.69	0.87	0.77	0.67

Tadqiqot doirasida yaratilgan algoritmlar asosida ishlab chiqilgan dasturiy ta'minot veb ilova ko'inishida yaratilgan bo'lib, mijoz-server arxitekturasini qo'llab quvvatlaydi. Dasturiy ta'minot tadqiqotchilar uchun mo'ljallangan bo'lib, unda tanlanmani yuklash, sozlamalarini to'g'irlash, turg'unlik natijalarini ko'rish, alomatlar juftligi bo'yicha mumkin bo'lmagan obyektlarni aniqlash, tanlanma uchun tasniflash aniqligini baholash, tanlanmani vizuallashtirish va yangi obyekttni tasniflash modullari mavjud.

XULOSA

“Berilganlarda noaniqlik bo'lganda qaror qabul qilish tizimi” mavzusidagi dissertatsiya ishi doirasida quyidagi asosiy natijalar olindi.

1. Berilganlarga dastlabki ishlov berish usullari, to'liq bo'lmagan berilganlarda bo'sh qiymatlari mavjud bo'lgan obyektlar ustida dastlabki ishlov berish usullari tahlil etildi. Tahlillar asosida tadqiqotning maqsadi va vazifalari shakllantirildi.

2. Miqdoriy alomatlar juftligi bo'yicha alomatning mumkin bo'lmagan qiymatlarini aniqlash usulining algoritmi ishlab chiqildi. Algoritm asosida tanlanmadagi anomal obyektlarni aniqlash imkoni paydo bo'ladi. Tizim ishlashi jarayonida ekspert yordamida bilimlar bazasini qurish algoritmi taklif qilindi.

3. To'liq bo'lmagan berilganlarga ishlov berishda muhim rol o'ynaydigan alomatlarining turg'unlik qiymatlari asoslandi. Turg'unlik qiymati bo'yicha alomatlar to'plamining informativligini baholash algoritmi va qaror qabul qilish tizimi ishlab chiqildi.

4. To'liq bo'lmagan berilganlarda alomatlar fazosi o'lchamini qisqartirish texnologiyasi taklif etildi. Latent alomatlar fazosida tasniflash algoritmlarini qo'llagan holda tasniflash aniqligi hisoblandi va taqqoslandi.

5. Alomatlar fazosini noxiziqli o'girish jarayonini amalga oshirishda parallel hisoblash usullaridan foydalanildi va samaradorliklari ko'rsatildi.

6. Yaratilgan algoritmlar asosida dasturiy ta'minotning funksional modellari ishlab chiqildi. Modelda miqdoriy alomatlar juftligi bo'yicha mumkin bo'lmagan obyekttni aniqlash, to'liq bo'lmagan berilganlarda alomatlar fazosini qisqartirish orqali bo'sh qiymatlarni yo'q qilish va tasniflash algoritmlari amalga oshirildi. Dasturiy ta'minot soha ekspertlari uchun qaror qabul qilishda yordamchi sifatida ishlatilib, yakuniy qaror ishonchligini baholash imkoniyatini yuzaga keltirdi.

**НАУЧНЫЙ СОВЕТ DSc.13/30.12.2021.Т.142.01
ПО ПРИСУЖДЕНИЮ УЧЁНЫХ СТЕПЕНЕЙ ПРИ НАУЧНО-
ИССЛЕДОВАТЕЛЬСКОМ ИНСТИТУТЕ РАЗВИТИЯ ЦИФРОВЫХ
ТЕХНОЛОГИЙ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ УЗБЕКИСТАНА

МАХАРОВ КОДИРБЕК ТОЛКУНОВИЧ

**СИСТЕМА ПРИНЯТИЯ РЕШЕНИЙ ПРИ НАЛИЧИИ
НЕОПРЕДЕЛЁННОСТИ В ДАННЫХ**

05.01.11 – Цифровые технологии и искусственный интеллект

**АВТОРЕФЕРАТ ДИССЕРТАЦИИ
ДОКТОРА ФИЛОСОФИИ (PhD) ПО ТЕХНИЧЕСКИМ НАУКАМ**

Ташкент – 2024

Тема диссертации доктора философии (PhD) по техническим наукам зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за номером B2023.4.PhD/T4194.

Диссертация выполнена в Национальном университете Узбекистана.

Автореферат диссертации на трех языках (узбекский, русский, английский (резюме)) размещен на веб-странице научного совета (www.airi.uz) и на Информационно-образовательном портале «ZiyoNet» (www.ziyo.net).

Научный руководитель: **Мадрахимов Шавкат Файзуллаевич**
доктор технических наук, профессор

Официальные оппоненты: **Мухамедиева Дилноз Тулкуновна**
доктор технических наук, профессор

Мирзаев Номаз
доктор технических наук, профессор

Ведущая организация: **Ташкентский государственный технический университет имени Ислама Каримова**

Защита диссертации состоится «06» 09 2024 года в 15⁰⁰ часов на заседании Научного совета DSc.13/30.12.2021.T.142.01 при Научно-исследовательском институте развития цифровых технологий и искусственного интеллекта. (Адрес: 100125, г. Ташкент, М.Улугбекский район, Буз-2, дом 17А. Тел.: (99871) 263-41-98; e-mail: info@airi.uz).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Научно-исследовательского института развития цифровых технологий и искусственного интеллекта (регистрационный номер № 45). (Адрес: 100125, г. Ташкент, М.Улугбекский район, Буз-2, дом 17А. Тел.: (99871) 263-41-98).

Автореферат диссертации разослан «23» 08 2024 года.
(протокол рассылки № 11 от «22» 07 2024 года)



Н. Равшанов

Председатель научного совета по
присуждению учёных степеней,
доктор технических наук, профессор

Ф.М. Нуралиев

Учёный секретарь научного совета
по присуждению учёных степеней,
доктор технических наук, профессор

Ш.Х. Фазилов

Председатель научного семинара при научном
совете по присуждению учёных степеней,
доктор технических наук, профессор

ВВЕДЕНИЕ (аннотация диссертации доктора философии (PhD))

Актуальность и востребованность темы диссертации. В связи с интенсивным развитием компьютерных технологий в мире и их применением в различных сферах человеческой деятельности наблюдаются потребность в качественной (интеллектуальной) обработке данных. Главным фактором инновационного развития общества являются знания. В развитых странах мира, таких как США, Великобритания, Германия, Франция, Российская Федерация, Япония, Китай, Южная Корея, в международных компаниях Amazon, Alibaba, Google, Yandex, Microsoft активно ведутся исследовательские работы по разработке и применению методов интеллектуального анализа данных в рамках систем искусственного интеллекта, ориентированных на решения задач прогнозирования, регрессии и классификации. Наряду с усвоением уже имеющихся знаний, актуальным является пополнение их объёмов за счёт открытия новых законов и закономерностей.

В настоящее время в мире широко исследуются методы предварительной обработки данных с целью построения на их основе информационных моделей, основанных на знаниях из слабо структурированных предметных областей. Ранняя очистка ложных данных стала серьезной проблемой при анализе данных в различных областях. Согласно анализу многих проведенных опросов, 38-39% всех данных имеют низкий показатель качества. Согласно анализу некачественных данных, 14-15% из них являются дубликатами, 9-11% – неверными значениями, 6-7% – неполными данными, а остальные – неправильными адресами электронной почты, несоответствие данных адресов проживания, ошибка формата и т.д. В качестве нерешенных проблем при формировании выборок данных для машинного обучения можно отметить отсутствие учета допустимости значений по парам признаков в описании объектов. Специфика этой проблемы связано с предметной областью, формированием латентных параметров для слежения за корректностью данных, созданием интеллектуальной системы для диагностики ошибок. Поэтому ранняя очистка неверных данных, обработка пропущенных данных и создание комплекса программ с удобным интерфейсом для широкого круга пользователей являются целенаправленными научными исследованиями.

Признавая большую значимость и потенциал технологий искусственного интеллекта в нашей стране, особое внимание уделяется разработке и внедрению систем на его основе. В частности, определены такие задачи, как “Установить, что руководители соответствующих министерств и ведомств, ответственных за формирование государственных данных для искусственного интеллекта, несут персональную ответственность за своевременное размещение информации в части, касающейся, а также их полноту, актуальность и достоверность”².

² Постановление Президента Республики Узбекистан от 18 февраля 2021 года ПП-4996 «О мерах по созданию условий для ускоренного внедрения технологий искусственного интеллекта»

Данное диссертационное исследование в определенной степени служит для выполнения задач, предусмотренных указом Президента Республики Узбекистан № УП-4947 от 7 февраля 2017 г. «О Стратегии действий по дальнейшему развитию Республики Узбекистан», № УП-5349 от 19 февраля 2018 года «О мерах по дальнейшему совершенствованию сферы информационных технологий и коммуникаций» и № УП-60 от 28 января 2022 года «О Стратегии развития Нового Узбекистана на 2022-2026 годы», постановлениями Президента Республики Узбекистан № ПП-3894 от 2 августа 2018 года «О мерах по внедрению инновационной модели управления здравоохранением в Республике Узбекистан», № ПП-4851 от 6 октября 2020 года «О мерах по дальнейшему совершенствованию системы образования в области информационных технологий, развитию и интеграции научных исследований с IT-индустрией», № ПП-4996 от 17 февраля 2021 года «О мерах по созданию условий для ускоренного внедрения технологий искусственного интеллекта», № ПП-357 от 22 августа 2022 года «О мерах по поднятию на новый уровень сферы информационно-коммуникационных технологий в 2022-2023 годах», а также другими нормативно-правовыми актами, принятыми в данной сфере.

Соответствие исследования приоритетным направлениям развития науки и технологий республики. Данное исследование выполнено в соответствии с приоритетным направлением развития науки и технологий Республики Узбекистан IV. «Информатизация и развитие информационно-коммуникационных технологий».

Степень изученности проблемы. В мире проведен ряд исследований по анализу неопределенностей в данных, выявлению скрытых закономерностей из базы данных, вычисление оценок объектов выборки и их практическому применению. Были сравнительно изучены научные работы Ю.И.Журавлева, В.А.Дюка, В.Н.Вапника, К.В.Воронцова, Т.А.Гавриловой, А.Зимека, Г.Литенса, Э.Рама, С.А.Захина, П.Дж.Гарсия-Лаенцина, В.Педрица, Дж.Лю и других зарубежных ученых. Ю.И.Журавлев разработал теорию алгебраического подхода к проблеме распознавания образов, в работах В.А.Дюка были предложены методы снижения комбинаторной сложности алгоритмов поиска логических закономерностей из данных, С.А.Захин, П.Дж.Гарсия-Лаенцина проводили исследования по решению задач классификации при неполных данных, В. Педриц предложил методы изолированного леса на основе алгоритмов ближайшего соседа для обнаружения аномальных объектов, Э. Рам исследовал проблемы очистки данных, качества данных и их классификации.

В Узбекистане можно указать ряд ученых, добившихся успехов в проблемах распознавания образов, обработки знаний в интеллектуальных системах, выявлении скрытых закономерностей, снижение размерности признакового пространства, определение эффективности полученных результатов, в том числе академик М.М.Камилов, Т.Ф.Бекмуратов, Ф.Т.Адилова, Ш.Х.Фазилов, Д.Т.Мухамадиева, Н.С.Маматов, Н.А.Игнатъев, М.М.Мусаев и др. внесли большой вклад в развитие этих направлений.

В реальных задачах встречаются различные неопределенности в данных, в частности, неполные данные и недопустимые значения в признаках. Исследования в этом направлении показывают, что интеллектуальные технологии (алгоритмы) принятия решений при неполных и логически несовместимых значениях заданных данных изучены недостаточно.

Связь диссертационного исследования с планами научно-исследовательских работ научно-исследовательского учреждения, где выполнена диссертация. Диссертационное исследование выполнено в рамках проектов плана научно-исследовательских работ Национального университета Узбекистана имени Мирзо Улугбека по теме ОТ-Атех-2018-224 “Разработка информационной модели прогнозирования развития, течения и исхода язвенного колита и колоректального рака, оптимизации диагностики и лечения с учетом этиопатогенеза и факторов риска” (2018-2021).

Целью исследования являются разработка алгоритмов и программного обеспечения (системы) для определения недопустимых значений в описании объектов выборки и решение задачи классификации с неполными данными.

Задачи исследования:

разработка алгоритма определения логического несоответствия в данных по парам количественных признаков в описании объектов;

разработка алгоритма оценки информативности совокупности признаков при неполных данных;

построение признакового пространства для классификации объектов на основе обобщенных оценок с учётом неполных данных;

применение парадигм параллельного программирования в процессе сокращения признакового пространство;

создание программного комплекса для решения практических задач на основе разработанных алгоритмов.

Объектом исследования являются информационные модели принятия решений на основе методов интеллектуальной обработки данных.

Предметом исследования являются алгоритмы поиска логического несоответствия в данных, вычисление границ интервалов по допустимым значениям по парам признаков, оценка устойчивости признаков, сокращение признакового пространства на основе обобщенной оценки при неполных данных.

Методы исследования. В ходе исследования применялись методы интеллектуальной обработки данных, машинного обучения, объектно-ориентированное программирование, параллельное программирование визуализации численных результатов.

Научная новизна исследования заключается в следующем:

разработан алгоритм определения логического несоответствия в данных по парам количественных признаков;

создана методика формирования базы знаний с учетом допустимых значений пар количественных признаков;

разработан алгоритм оценки информативности наборов признаков и система принятия решений на основе свойства (меры) их устойчивости;

разработан алгоритм снижения размерности признакового пространства с использованием технологии вычисления обобщенных оценок с учетом пропусков в данных.

Практические результаты исследования заключаются в следующем:

предложена методология поддержки принятия решений по аномальным объектам предметной области;

предложено решение проблемы классификации по неполным данным путем формирования нового, сокращенного пространства признаков;

разработана IDEF модель принятия решений на основе предварительной обработки при наличии неопределенности в данных;

предложен параллельный алгоритм для достижения эффективности по времени при формировании пространства латентных признаков.

Достоверность результатов исследования. Достоверность результатов исследования обосновывается при проведении вычислительных экспериментов на выборках данных из предметных областей, на базах данных из репозиториях, размещенных на специализированных сайтах глобальной сети Internet, использованием критериев устойчивости, сравнением с результатами других авторов.

Научная и практическая значимость результатов исследования.

Научная значимость результатов исследования выражается в разработке методики принятия решений, основанной на определении недопустимых значений по парам количественных признаков при описании объектов и снижении размерности признакового пространства с использованием обобщенных оценок.

Практическая значимость результатов исследования заключается в том, что разработанные алгоритмы и созданное на их основе программное обеспечение могут быть использованы для решения задач классификации по неполным данным в различных предметных областях, для определения логически недопустимых значений пары признаков, а также созданием программных средств для поддержки принятия решений экспертами предметных областей и их применением к конкретным задачам.

Внедрение результатов исследования. На основе научных результатов, полученных по разработанным алгоритмам и их программного обеспечения, для принятия решений при наличии неопределенности в данных:

алгоритм вычисления интервалов допустимых значений по парам количественных признаков, алгоритм оценки достоверности принимаемых решений при наличии неопределенности в данных был внедрен в Научно-практическом центре интеллектуальных программных систем при Национальном университете Узбекистана им. Мирзо Улугбека (справка Министерства цифровых технологий Республики Узбекистан от 16 мая 2024 года, № 33-8/3190). В результате в процессе формирования базы данных по туберкулезу легких удалось сократить время проверки данных и принятия решений на 10-15%;

система принятия решений, созданное на основе алгоритма сокращения признакового пространства с использованием метода вычисления

обобщенных оценок с целью исключения случаев неопределенности в данных и параллельного алгоритма процесса формирования латентных признаков была применена в ГУП “Центр научно-технических и маркетинговых исследований UNICON.UZ” (справка Министерства цифровых технологий Республики Узбекистан от 16 мая 2024 года, № 33-8/3190). В результате удалось сократить время обработки данных и процесса принятия решений на 15-20%.

Апробация результатов исследования. Результаты данного исследования были обсуждены на 15 международных и 6 республиканских научно-практических конференциях.

Опубликованность результатов исследования. По теме диссертации опубликовано 35 научных работ, в том числе, 9 журнальных статей, 4 в республиканских и 5 в иностранных научных изданиях, рекомендованных Высшей аттестационной комиссией Республики Узбекистан для публикации основных научных результатов диссертаций. Получены 5 свидетельств об официальной регистрации программ для ЭВМ.

Структура и объем диссертации. Диссертация состоит из введения, четырех глав, заключения, списка использованной литературы и приложений. Объем диссертации составляет 106 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснованы актуальность и востребованность темы диссертации, показано соответствие исследования приоритетным направлениям развития науки и технологий Республики Узбекистан. Определены цели и задачи исследования, а также указаны объект и предмет исследования, изложены научная новизна и практические результаты исследования, обоснована достоверность полученных результатов, раскрыта теоретическая и практическая значимость полученных результатов, приведены перечень внедрений результатов исследования в практику, сведения об опубликованных работах и о структуре диссертации.

В первой главе диссертации под названием “**Предварительная обработка данных**” был представлен информации о данных и их свойств, формирования пространства признаков, сравнительный анализ методов и алгоритмов первичной обработки данных. В частности, представлены средства и виды хранения данных, шкалы возможных значений признаков. В настоящее время данные генерируются практически везде: региональные сети сейсмических датчиков, анкетирование и учет состояния больных в медицинских учреждениях, опросы общественного мнения по определенной теме, база данных активностей клиентов в провайдерах мобильной связи и т.д. Большая часть данных из реального мира имеет общий недостаток – отсутствие или неизвестность данных. Например, в промышленных экспериментах некоторые результаты могут быть недоступны из-за механических/электронных сбоев во время сбора данных, в медицинской диагностике некоторые тесты невозможно выполнить, поскольку в больнице нет необходимого медицинского оборудования или какое-то медицинское

оборудование отсутствует, некоторые тесты могут не подойти некоторым пациентам. Другим примером может быть осмотр врачом, который проводит различные виды анализов; результаты некоторых тестов могут быть доступны сразу, а другие могут занять несколько дней. В любом случае может возникнуть необходимость немедленно поставить предварительный диагноз, используя только имеющиеся результаты. Для этого используются три разных подхода: удаление объектов или признаков с пропущенными значениями; заполнение пропущенных значений; решить задачу с пропущенными значениями. Первые два подхода имеют ряд недостатков, поэтому одной из актуальных задач является разработка методов и алгоритмов решения задач с пропущенными значениями.

Во второй главе диссертации **“Алгоритм контроля допустимых значений признаков”** предлагается метод обнаружения логически недопустимых значений по парам количественных признаков в описании объектов выборки в исследуемой предметной области и алгоритм ориентированный на обнаружение ошибок во входных данных.

В виде таблицы $A = \{a_{uv}\}_{m \times n}$ задана выборка из множества объектов $E_0 = \{S_1, \dots, S_m\}$ описываемых набором количественных признаков $X = (x_1, \dots, x_n)$. По каждой паре признаков $(x_i, x_j) \in X(n)$ требуется определить:

- границы допустимых интервалов $[z_1, z_2]$, по

$$R(i, j) = \left\{ \frac{a_{ki}}{P_i} - \frac{a_{kj}}{P_j} \right\}_{k \in (1, \dots, m)}, \quad (1)$$

где $S_k = (a_{k1}, \dots, a_{kn})$, $S_k \in E_0$, P_i, P_j – значения медиан признаков x_i и x_j на множестве E_0 ;

- значения границ интервалов z_1 и z_2 :

$$z_1 = \min_{E_0} R(i, j), z_2 = \max_{E_0} R(i, j). \quad (2)$$

Следует отметить, что только специалист предметной области может дать оценку о логической правильности значений пары признаков. Этот метод может быть реализован путем привлечения соответствующего эксперта к первоначальной обработке следующим образом:

1. Для каждой пары признаков (x_i, x_j) , $i < j$, $1 \leq i, j \leq n$ построение множество интервалов $Z = \{[z_1^j, z_2^j]\}$;
2. Построение множество объектов, лежащих на границе интервала:

$$\Pi = \bigcup_{i, j \in I, i \neq j} \left\{ S_u, S_v \left| \frac{a_{ui}}{P_i} - \frac{a_{uj}}{P_j} = z_1 \text{ и } \frac{a_{vi}}{P_i} - \frac{a_{vj}}{P_j} = z_2 \right. \right\};$$

3. Для объекта $\forall S_u \in \Pi$ принять мнение эксперта о допустимости значений (оставить без изменений, исправить заданные значения или удалить объект);
4. Пересчет границ интервала допустимых значений по (1) и (2) при изменении выборки.

Сформированный алгоритмом база интервалов Z может использоваться для проверки допустимости значений в описании нового объекта при решении поставленной задачи. Можно различить два типа ошибок, которые могут возникнуть при определении недопустимого объекта:

- “*критическая ошибка*” – для объекта, значение которого по паре признаков лежит на границах или за границами интервала;

- “*предупреждение*” – для объекта, значение которого по паре признаков “*опасно*” близко к границам интервала.

Для вычислительного эксперимента использована демографическая база собранным Nancy Howell - “Kalahari !Kung San people”. Для эксперимента использовались только признаки “*возраст*”, “*рост*” и “*вес*”. В 1-таблице представлены результаты значений интервалов по парам признаков.

1-таблица.

Значения интервалов по парам признаков

	Пара признаков	Интервал [z ₁ ,z ₂]	Объекты на границе интервала (значения пары признаков)	
			слева	справа
1	(Рост, Вес)	[-0.58, 0.46]	5 (163.83, 62.99)	359 (109.22, 11.71)
2	(Рост, Возраст)	[-1.90, 0.72]	164 (140.97, 85.60)	359 (109.22, 2.00)
3	(Вес, Возраст)	[-1.79, 0.99]	222 (35.81, 82.00)	182 (61.80, 22.00)

В табл. 2 приводятся примеры результатов проверки на наличие недопустимых объектов с указанием статуса ошибок: значения признаков по (2) лежат за границами интервала (критическая ошибка); значения признаков по (2) лежат в “опасной” близости от границ интервала (предупреждение).

2-таблица.

Представление потенциально недопустимых объектов

Номер объекта	Название признака	Значение признаков	Интервал	Значение (2) по паре признаков	Статус ошибки
6	Рост	163.83	[-0.41, 0.44]	-0.47	<i>Критическая ошибка</i>
	Вес	62.99			
90	Рост	136.53	[-2.23, 0.66]	-2.01	<i>Предупреждение</i>
	Возраст	79.00			
90	Вес	36.29	[-2.15, 0.73]	-2.02	<i>Предупреждение</i>
	Возраст	79.00			
165	Вес	40.94	[-2.14, 0.73]	-2.15	<i>Критическая ошибка</i>
	Возраст	85.60			
174	Вес	32.21	[-2.15, 0.73]	-2.12	<i>Предупреждение</i>
	Возраст	79.00			

На рис. 1 показано расположение объектов выборки по парам признаков между границ допустимых интервалов по предложенному методу. Символом “X” обозначены объекты (аномальные объекты), у которых может быть

логически недопустимые значения. Аномальными объектами считаются объекты, расположенные на $5h$ расстоянии от границы интервала, $h = (z_2 - z_1) / 100$.

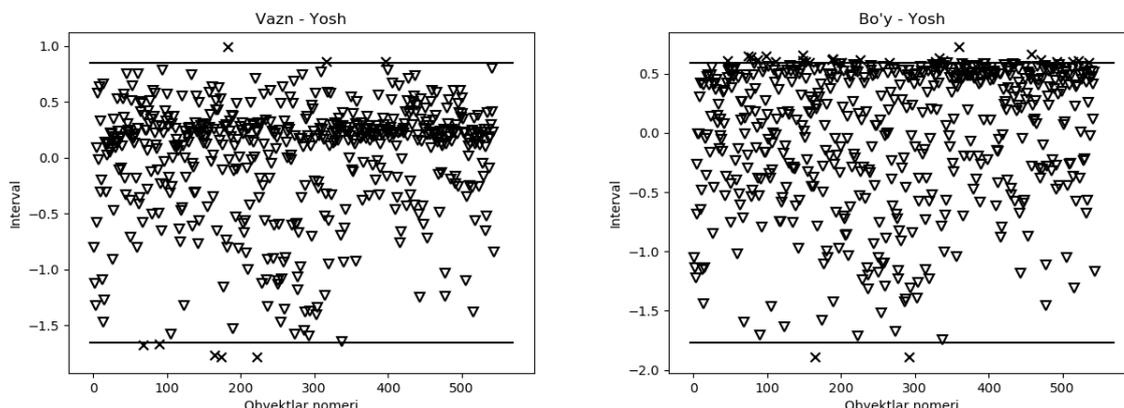


Рисунок 1. Графическое представление объектов по парам признаков

Для сравнения методов обнаружения аномальных объектов был проведен эксперимент с искусственным добавлением 20 аномальных объектов к исходным данным. Поскольку предложенный метод работает с парой признаков, в ходе эксперимента объект считался аномальным если по одной паре признаков есть недопустимые значения. Для обеспечения более точного сравнения использовались точность (precision) и оценка f_1 (f_1 -score). Результаты сравнения представлены в табл. 3.

3-таблица.

Сравнительные результаты по обнаружению аномальных объектов

Тест №	Isolation Forest		DBSCAN		Предлагаемый метод	
	оценка f_1	точность	оценка f_1	точность	оценка f_1	точность
Тест 1	0.70	0.70	0.92	0.90	0.85	1.00
Тест 2	0.80	0.80	0.90	0.86	0.91	1.00
Тест 3	0.75	0.75	0.90	0.86	0.88	1.00
Тест 4	0.70	0.70	0.90	0.86	0.91	1.00
Тест 5	0.70	0.70	0.92	0.92	0.72	0.62
Тест 6	0.65	0.65	0.90	0.90	0.88	1.00
Тест 7	0.60	0.60	0.90	0.90	0.91	1.00
Тест 8	0.70	0.70	0.88	0.82	0.82	0.77
Тест 9	0.75	0.75	0.88	0.82	0.91	1.00
Тест 10	0.70	0.70	0.85	0.81	0.88	1.00
Среднее	0.70	0.70	0.89	0.86	0.87	0.94

Согласно табл. 3, в результатах метода DBSCAN оценка f_1 превосходил предлагаемый метод во многих тестах, но точность рекомендуемого метода выше. Фактически, поскольку задача состоит в поиске аномальных объектов, можно сделать вывод, что предлагаемым методом можно найти больше аномальных объектов исходя от значения точности (доля правильно найденных положительных значений).

В третьей главе диссертации “Алгоритмы принятия решений и оценка их достоверности на неполных данных” обоснована важность

свойства устойчивости признаков, предложен метод оценки информативности набора признаков, сокращения пространство признаков на основе обобщённых оценок, представлены результаты применения алгоритмов классификации и оценки точности классификации.

Постановка задачи. Пусть задано множество объектов $E_0 = \{S_1, \dots, S_m\}$, содержащее представителей непересекающихся классов K_1 и K_2 . Описание объектов производится с помощью набора из n разнотипных признаков $X(n) = \{x_1, \dots, x_n\}$, δ ($\delta < n$) из которых измеряются в номинальной, $n - \delta$ в количественной шкале. Допускаются наличие пропусков в данных. Одной из причин рассмотрения проблемы двух классов является то, что обобщенное значение любого объекта относительно, оно возникает в результате сравнения с объектами противоположного класса. Во-вторых, любую задачу класса k ($k > 2$) можно решить в виде каскада двухклассовых задач. Пусть $\mu(x_i)$ – неотрицательная мера, определенная в наборе измеренных значений признаков, относящихся к $X(n)$. Требуется:

- вычислить границы интервалов доминирования для количественных признаков;
- вычислить значения функции принадлежности интервалов (для номинальных признаков каждой градации) к классу K_1 ;
- оценить разброс значений $\mu(x_i)$ в зависимости от числа измеренных значений;
- вычислить информативность набора признаков на основе значений $\mu(x_i)$;
- произвести нелинейное преобразование признакового пространства на основе функций принадлежности;
- произвести выбор латентного пространства путем вычисления обобщённых оценок объектов по указанным наборам признаков;
- оценить точность классификации в новом признаковом пространстве.

Чтобы доказать объективность оценки значения устойчивости, был проведен эксперимент со случайным распределением пропусков в описании объектов. Процент пропущенных значений по каждому признаку варьировался от 0,0% до 35,0%. Результаты эксперимента по изучению влияния пропущенных значений на устойчивость признаков представлены в табл. 4. Анализ значений дисперсии в таблице позволяет сделать вывод о низкой вариабельности (изменчивости) значения устойчивости даже в случае наличия пропусков в значениях.

4-таблица.

Результаты эксперимента по значениям устойчивости

Признак	Мат. Ожидание	Дисперсия	Признак	Мат. Ожидание	Дисперсия
Region	0.6016	0.000068	Normal Chair	0.5339	0.000250
Occupation	0.6094	0.000105	Constipation	0.5566	0.000064
Age	0.6204	0.000174	Stool Frequency	0.5351	0.000176
Gender	0.5175	0.000244	Chair Character	0.5380	0.000100
Pyrazinamide	0.5464	0.000052	Eosinophils	0.5862	0.000124
Ethambutol	0.5111	0.000074	Lymphocytes	0.6015	0.000760
...

Обозначим через

$$\omega_1(t_1), \dots, \omega_g(t_g), \dots, \omega_n(t_n), \quad t_i \in \{1, \dots, n\} \quad (3)$$

последовательность значений математических ожиданий признаков, упорядоченных в порядке невозрастания. Индекс элементов в (3) интерпретируется как ранг признака. Определим информативность набора из k ($[n/2] < k < n$) признаков без пропусков в описании произвольного допустимого объекта S . На основании значений табл. 4, упорядоченная последовательность признаков выглядит следующим образом:

$$\begin{aligned} & x_{48}, x_{14}, x_{16}, x_{15}, x_{40}, x_{36}, x_{38}, x_1, x_3, x_{32}, x_{39}, x_2, x_{47}, x_{43}, x_{45}, x_{44}, \\ & x_{46}, x_{18}, x_5, x_{42}, x_{17}, x_9, x_{26}, x_{20}, x_{23}, x_{25}, x_{24}, x_8, x_{33}, x_{34}, x_{27}, x_{19}, \\ & x_{22}, x_{28}, x_{29}, x_{35}, x_{41}, x_{12}, x_{37}, x_6, x_{31}, x_{21}, x_{13}, x_{30}, x_{11}, x_4, x_{10}, x_7. \end{aligned} \quad (4)$$

Пусть $D(S, k)$ – сумма рангов k измеренных значений признаков объекта S . Значение суммы находится между $k(1+k)/2 \leq D(S, k) \leq k(2n-k+1)/2$. Для оценки информативности набора из k признаков объекта S предлагается использовать меру

$$\Omega(S, k) = 1 - \left(\frac{D(S, k) - \alpha}{\beta - \alpha} \right) \left(1 - \frac{k}{n} \right), \quad (5)$$

где $\alpha = k(1+k)/2$, $\beta = k(2n-k+1)/2$. Применение меры (5) рекомендуется при анализе данных в описании произвольных допустимых объектов. Например, его можно использовать как дополнительный показатель при принятии решения по объекту с пропущенными значениями (меру достоверности решения). В табл. 5 приведены примеры применения меры информативности по некоторым наборам признаков согласно (4). Диапазон возможных значений меры информативности в заключении эксперт может определить самостоятельно. В примере в таблице диапазоны определены следующим образом:

- $[0, 0,6)$ – “недостаточно”;
- $[0,6, 0,72)$ – “удовлетворительно”;
- $[0,72, 0,86)$ – “достаточно”;
- $[0,86, 1,0]$ – “соответствует”.

5-таблица.

Мера информативности для комбинаций признаков

№	Номера используемых признаков	Вывод	Значение (5)
1	4, 6, 8, 9, 10, 11, 13, 17, 19, 21, 23, 24, 25, 27, 29, 30, 31, 33, 35, 41, 42	<i>недостаточно</i>	0,5179
2	4, 6, 7, 8, 10, 11, 13, 19, 21, 22, 27, 28, 29, 30, 31, 33, 34, 35, 37, 41	<i>недостаточно</i>	0.4271
3	8, 9, 12, 17, 19, 20, 22, 23, 26, 27, 29, 33, 34, 35, 36, 37, 39, 42, 44, 46	<i>удовлетворительно</i>	0.6813
4	2, 5, 6, 8, 13, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 35, 37, 39, 41, 45	<i>удовлетворительно</i>	0.6313
5	1, 5, 8, 9, 17, 19, 23, 25, 26, 27, 28, 32, 33, 35, 37, 38, 40, 42, 43, 44, 46	<i>достаточно</i>	0.7708
6	2, 3, 5, 9, 12, 17, 18, 19, 20, 22, 25, 26, 28, 33, 35, 36, 38, 40, 43, 44, 46, 47	<i>достаточно</i>	0.8097

№	Номера используемых признаков	Вывод	Значение (5)
7	1, 2, 3, 5, 9, 14, 15, 17, 18, 19, 22, 23, 24, 26, 32, 33, 34, 38, 39, 42, 46, 48	<i>соответствует</i>	0.8797
8	3,5, 9, 14, 15, 16, 17, 20, 26, 32, 34, 38, 39, 40, 42, 43, 44, 46, 47, 48	<i>соответствует</i>	0.9365

Суть нелинейных преобразований признаков заключается в замене их исходных значений на значения функции принадлежности к классам. Для $x_c \in X(n)$ граница между объектами классов определяется как:

$$G_c = \frac{q_1 + q_2}{2}, \quad (6)$$

где $q_1 = \min\{f_c(\gamma) \mid 1 - f_c(\gamma) < 0.5, \gamma = 1, \dots, \tau_c\}$,
 $q_2 = \max\{f_c(\gamma) \mid 0.5 - f_c(\gamma) > 0, \gamma = 1, \dots, \tau_c\}$.

При вычислении значение градации α_{ic} по (6) для объекта $S_i = \{x_{iu}\}_{u \in D}$ рассматривается одно из двух условий с учетом масштабов измерений: $x_{ic} \in [r_u, r_v]^v$ или $x_{ic} = \gamma$. Для вычисления значения α_{ic} используется функция принадлежности:

$$a_{ic} = \begin{cases} 1, & f_c(\gamma) < G_c, \\ 2, & f_c(\gamma) > G_c. \end{cases}$$

Для вычисления обобщённых оценок объектов используется вклады признаков, описываемые градациями из $\{1, 2\}$. Вклад признака x_c по градации $j \in \{1, 2\}$ определяется как

$$\eta_c(j) = w_c \left(\frac{\alpha_{cj}^1}{T_{1c}} - \frac{\alpha_{cj}^2}{T_{2c}} \right), \quad (7)$$

где $\alpha_{cj}^1, \alpha_{cj}^2$ – количество значений градации j признака x_c соответственно в классах K_1 и K_2 , w_c – вес признака x_c , вычисляемый как произведение внутриклассового сходства и межклассовых различий. Обобщённая оценка объекта $S_r \in E_0$ по описанию в номинальной шкале измерений вычисляется как:

$$R(S_r) = \sum_{i \in D} \eta_c(\alpha_{ri}). \quad (8)$$

При построении нового признакового пространства в качестве основного критерия принимается значение устойчивости. Значение устойчивости признака попадают в интервал $(0.5; 1]$. Если значение устойчивости равно единице, объекты выборки разделяются на классы K_1 и K_2 без ошибок. Удаление признаков с относительно небольшими значениями устойчивости предлагается для снижения вычислительной сложности в алгоритмах анализа данных.

Пусть Ω , $\Omega \subset X(n)$ обозначает набор признаков с относительно малым значением устойчивости, $\Pi = \{i \mid i \in X(n) \setminus \Omega\}$ – индексы признаков, используемых в группировке. При формировании латентных признаков

используются разные группы признаков из Π . Признаки сортируются по значению устойчивости и группируются по k ($k = 2, \dots, |\Pi|/2$). Оптимальное значение k определяется вычислительным экспериментом. Для каждой группы признаков вычисляются обобщенные оценки по (7) и (8). Значение обобщенных оценок используется как значения латентных признаков в описании объекта в новом признаковом пространстве без пропусков.

В первом вычислительном эксперименте были использованы две выборки. Первый из них данные о больных туберкулезом, представленные 167 объектами, описываемые 48 признаками (16 из них количественными, 32 – номинальными). Второй – набор данных о раке молочной железы в Висконсине, представленный 569 объектами, описываемый 30 признаками (все количественные). Пустые значения от 0,0% до 50,0% были искусственно внедрены для каждого признака. Оптимальное значение k определялось вычислительным экспериментом. Точность классификации в новом признаковом пространстве была проверена с использованием следующих методов машинного обучения:

- Байесовский классификатор (Naïve Bayes);
- Метод k-ближайших соседей (k-nearest neighbors, KNN);
- Метод опорных векторов (Support Vector Machine, SVM);
- Случайный лес (Random Forest);
- Дерево решений (Decision Tree);
- Многослойный перцептрон (Multi-layer Perceptron, MLP).

В табл. 6 приведены результаты точности классификации, на выборках с искусственно сгенерированными пустыми значениями от 0,0% до 50,0% для каждого признака (при $k=5$). Для определения результатов точности классификации было проведено десять экспериментов и получено их среднее значение. Результаты показывают, что даже при наличии пропусков (до 50%) в новом признаковом пространстве достигаются результаты высокой точности классификации.

6-таблица.

Результаты эксперимента по исследованию точности классификации

№	Пропуски (%)	Точность классификатора					
		Naïve Bayes	KNN	SVM	Random Forest	Decision Tree	MLP
Больные туберкулезом							
1	0	0.960	0.960	0.940	0.980	0.920	0.980
2	5	0.958	0.958	0.952	0.952	0.914	0.958
3	10	0.956	0.956	0.936	0.956	0.912	0.948
4	15	0.936	0.906	0.932	0.932	0.884	0.930
5	20	0.926	0.948	0.910	0.934	0.890	0.926
6	25	0.946	0.926	0.924	0.934	0.890	0.932
7	30	0.916	0.930	0.936	0.942	0.894	0.920
8	35	0.940	0.940	0.920	0.942	0.888	0.924
9	40	0.946	0.944	0.922	0.948	0.902	0.914
10	45	0.924	0.940	0.914	0.938	0.880	0.906
11	50	0.928	0.922	0.916	0.930	0.890	0.912

№	Пропуски (%)	Точность классификатора					
		Naïve Bayes	KNN	SVM	Random Forest	Decision Tree	MLP
Больные раком молочной железы							
1	0	0.976	0.977	0.977	0.977	0.942	0.959
2	5	0.970	0.966	0.971	0.968	0.945	0.958
3	10	0.966	0.967	0.971	0.965	0.952	0.956
4	15	0.969	0.959	0.965	0.967	0.946	0.958
5	20	0.968	0.961	0.965	0.964	0.934	0.954
6	25	0.966	0.964	0.969	0.963	0.950	0.946
7	30	0.956	0.960	0.955	0.961	0.940	0.944
8	35	0.963	0.954	0.958	0.957	0.923	0.950
9	40	0.960	0.954	0.963	0.960	0.929	0.949
10	45	0.964	0.957	0.967	0.960	0.936	0.948
11	50	0.958	0.953	0.964	0.950	0.931	0.949

Во втором вычислительном эксперименте предложенный метод сравнивался с другими подходами. Для этой цели была использована выборка “Диабет у индейцев Пима”. Выборка описывается 768 объектами и 8 признаками. 500 (65,1%) объектов относятся к классу здоровых и 268 (34,9%) к классу больных. В пяти признаков выборки (глюкоза, артериальное давление, толщина кожи, инсулин и индекс массы тела) имеются значения равные нулю, но в реальности эти признаки не могут принимать значение ноль. Следовательно, эти значения можно считать ошибочными (пропусками). В результате количество пропущенных значений распределилось по признакам следующим образом:

- глюкоза: 5 (0.65%);
- артериальное давление: 35 (4.56%);
- толщина кожи: 227 (29.56%);
- инсулин: 374 (48.70%);
- индекс массы тела: 11 (1.43%).

Всего у 376 объектов имеются пропуски в значениях. Это означает 48,96 процента от всей выборки. В табл. 7 сравнивается подход к устранению пропущенных значений предложенный в этом исследовании с другими полученными результатами и подходами. Точность результатов алгоритмов классификации с использованием предложенного метода сокращения пространства во многих случаях дала хорошие результаты. Наилучший результат был достигнут, когда параметр группировки признаков – $k=3$ и для классификации использовался наивный байесовский алгоритм.

7-таблица.

Результаты эксперимента при данных с пропусками

№	Используемый подход	Точность классификации					
		Naïve Bayes	KNN	SVM	Random Forest	Decision Tree	MLP
1.	Удаление неполных объектов	0.738	0.720	0.729	0.771	0.695	0.754
2.	Заполнение средним значением	0.740	0.709	0.745	0.745	0.692	0.745

3.	Заполнение методом к ближайших соседей (k=11)	0.736	0.701	0.727	0.757	0.732	0.746
4.	Нейронная сеть общей регрессии	0.802					
5.	Метод градиентного бустинга	0.789					
6.	Метод Logdisc	0.777					
7.	Метод IncNet	0.776					
8.	Предлагаемый метод: k=2	0.761	0.799	0.785	0.777	0.757	0.788
	k=3	0.814	0.758	0.788	0.766	0.746	0.775
	k=4	0.809	0.788	0.771	0.771	0.766	0.796

В четвертой главе диссертации “Архитектура программного обеспечения и ее применение при решении практических задач” исследуются архитектуры программного обеспечения, используемые при создании современных интеллектуальных информационных систем, интеллектуальных систем поддержки принятия решений и предлагается использование микросервисной гибридной архитектуры, а также перечислено какие микросервисы должно содержать созданное программное обеспечение. Предложены функциональные модели создаваемого программного обеспечения, посредством вычислительного эксперимента показана эффективность использования принципов параллелизма в программном обеспечении, показан интерфейс созданного программного обеспечения и пример его использования в практической задаче. Созданы IDEF0 функциональная (рис. 2) и IDEF1.x информационная модели создаваемого программного обеспечения по предложенным алгоритмам, а также схема работы системы, которая может быть создана для медицинской сферы на основе функциональных моделей.

В данной исследовательской работе этапы определения интервалов доминирования количественных признаков и вычисления значения функции принадлежности интервалов к классу при построении нового (латентного) признакового пространства имеют вычислительную сложность, связанную с полным перебора, и при реализации этих методов использование архитектуры Multiple Instruction Multiple Data приводит к высокой эффективности. Проведен вычислительный эксперимент по оценке эффективности использования принципов параллелизма при использовании алгоритмов. Для вычислительного эксперимента была выбрана выборка для задачи диагностики рака простаты. Эта выборка состоит из 102 объектов и 12625 признаков. Эксперимент проводился на компьютере с процессором Intel(R) Core(TM) i7-8550U, 1,80 ГГц (объем оперативной памяти – 8Gb, объем графической памяти – 2Gb). Процессор имеет 4 физических ядра и 8 логических ядер. В табл. 8 представлены результаты использования параллельных вычислений при реализации процесса нелинейного преобразования признакового пространства. В эксперименте количество

признаков было искусственно увеличено с целью увеличения размера выборки. Эффективность использования параллелизма приведена по закону Амдала (табл. 9, доля программы, которую можно получить только путем последовательных вычислений = 0.125).

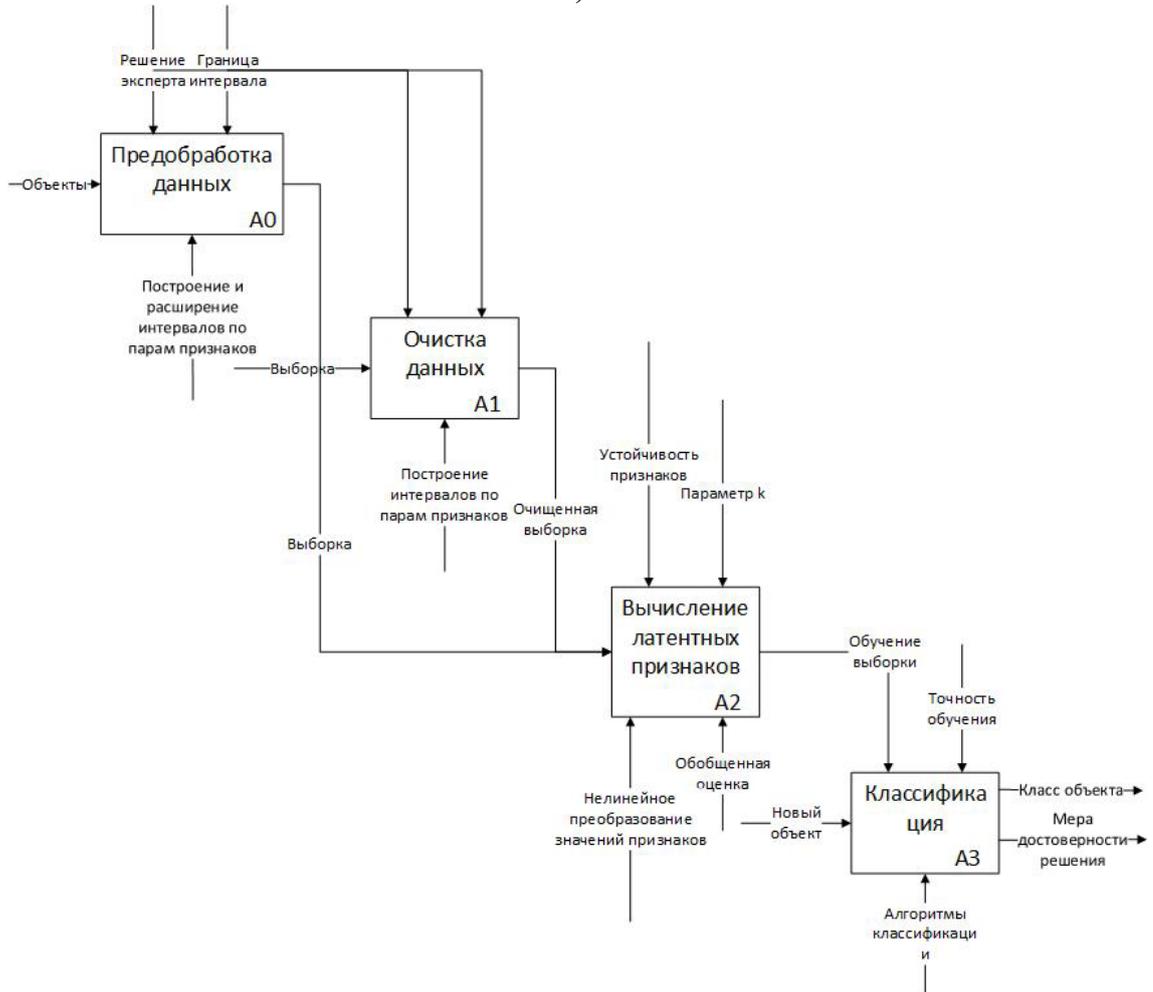


Рисунок 2. IDEF0 функциональная модель системы

8-таблица.

Результаты параллельных вычислений

Количество признаков	Количество ядер процессора							
	1	2	3	4	5	6	7	8
12625	17.02	10.58	9.14	8.10	8.32	8.78	9.48	12.17
126250	158.71	97.56	73.18	65.19	61.14	65.79	63.79	68.01
252500	335.05	184.49	143.94	126.38	135.99	129.01	117.66	122.93
378750	494.55	291.08	227.54	198.09	193.54	188.15	182.27	183.58
505000	689.10	385.37	298.90	268.16	265.99	288.04	292.64	278.94

9-таблица.

Эффективность параллельных вычислений

Количество приз. / Кол. ядер	Ускорение			Эффективность		
	2	3	4	2	3	4
Идеальная	1.78	2.40	2.91	0.89	0.80	0.73
12625	1.61	1.86	2.10	0.80	0.62	0.53

126250	1.63	2.17	2.43	0.81	0.72	0.61
252500	1.70	2.18	2.57	0.85	0.73	0.64
378750	1.70	2.23	2.60	0.85	0.74	0.65
505000	1.74	2.31	2.69	0.87	0.77	0.67

Программное обеспечение, разработанное на основе алгоритмов, созданных в рамках исследования, создано в виде веб-приложения и поддерживает архитектуру клиент-сервер. Программное обеспечение предназначено для исследователей и включает в себя модули загрузки выборки, настройки его параметров, просмотра результатов устойчивости признаков, идентификации недопустимых объектов по паре признаков, оценки точности классификации в выборке, визуализации выборки и классификации нового объекта.

ЗАКЛЮЧЕНИЕ

В рамках диссертационной работы на тему “Система принятия решений при наличии неопределенности в данных” получены следующие основные результаты.

1. Проанализированы методы первичной обработки данных, методы предварительной обработки объектов с пустыми значениями в неполных данных. На основе анализа были сформированы цель и задачи исследования.

2. Разработан алгоритм метода определения недопустимых значений признака по паре количественных признаков. На основе алгоритма становится возможным определять аномальные объекты в выборке. Предложен алгоритм построения базы знаний с помощью экспертов в процессе работы системы.

3. Обосновано значения стабильности признаков, играющие важную роль при обработке неполных данных. Разработан алгоритм оценки информативности множеств признаков и система принятия решений на основе значений устойчивости.

4. Предложена технология снижения размерности признакового пространства при неполных данных. Была вычислена и сравнена точность классификации с использованием алгоритмов классификации в латентном признаковом пространстве.

5. При реализации процесса нелинейного преобразования признакового пространства были использованы методы параллельных вычислений и показана их эффективность.

6. На основе созданных алгоритмов разработаны функциональные модели программного обеспечения. В моделях реализованы алгоритмы определения недопустимого объекта по паре количественных признаков, исключения пустых значений путем сокращения признакового пространства в неполных данных и классификация. Программное обеспечение использовалось в качестве помощника при принятии решений для экспертов предметной области, позволяет оценить достоверность окончательного решения.

**SCIENTIFIC COUNCIL AWARDING SCIENTIFIC DEGREES
DSc.13/30.12.2021.T.142.01 AT RESEARCH INSTITUTE FOR
DEVELOPMENT OF DIGITAL TECHNOLOGIES AND ARTIFICIAL
INTELLIGENCE**

NATIONAL UNIVERSITY OF UZBEKISTAN

MAKHAROV KODIRBEK TOLKUNOVICH

**DECISION MAKING SYSTEM IN THE PRESENCE OF
UNCERTAINTY IN DATA**

05.01.11 – Digital technologies and artificial intelligence

**DISSERTATION ABSTRACT OF THE DOCTOR OF PHILOSOPHY (PhD)
ON TECHNICAL SCIENCES**

Tashkent – 2024

The theme of doctor of philosophy (PhD) on technical sciences was registered at the Supreme Attestation Commission at the Cabinet of Ministers of the Republic of Uzbekistan under number B2023.4.PhD/T4194.

The dissertation has been prepared at National university of Uzbekistan.

The abstract of the dissertation is posted in three languages (uzbek, russian, english (resume)) on the website (www.airi.uz) and on the website of "Ziyonet" Information and educational portal (www.ziyonet.uz).

Scientific adviser: **Madrakhimov Shavkat Fayzullaevich**
Doctor of Technical Sciences, professor

Official opponents: **Mukhamediyeva Dilnoz Tulkunovna**
Doctor of Technical Sciences, professor

Mirzaev Nomaz
Doctor of Technical Sciences, professor

Leading organization: **Tashkent State Technical University**
named after Islam Karimov

The defense will take place "06" 09 2024 at 15⁰⁰ on the meeting of Scientific council No. DSc.13/30.12.2021.T.142.01 at Research Institute for Development of Digital Technologies and Artificial Intelligence (Address: 17A, Buz-2, M.Ulugbek district, Tashkent city, 100125/ Phone: (+99871) 263-41-98, e-mail: info@airi.uz).

The dissertation is available at the Information Resource Centre of the Research Institute for Development of Digital Technologies and Artificial Intelligence (is registered under No. 45). (Address: 17A, Buz-2, M.Ulugbek district, Tashkent city, 100125/ Phone: (+99871) 263-41-98).

Abstract of dissertation sent out on "23" 08 2024 y.
(mailing report No. 11 on "22" 07 2024 y.)



N.Ravshanov
Chairman of the scientific council
awarding scientific degrees,
Doctor of Technical Sciences, Professor

F.M.Nuraliev
Scientific secretary of scientific council
awarding scientific degrees,
Doctor of Technical Sciences, Professor

Sh.Kh.Fazilov
Chairman of the academic seminar under
the scientific council awarding scientific degrees,
Doctor of Technical Sciences, Professor

INTRODUCTION (abstract of PhD dissertation)

The aim of the research work is to develop algorithms and software (system) for determining invalid values in the description of sample objects and solving the classification problem with incomplete data.

The object of the research work is information model that make decisions based on data mining methods.

The scientific novelty of the research work is consist on follows:

developed an algorithm for determining the logical inconsistencies of a pair of quantitative features;

developed a methodology for forming a knowledge base taking into account allowable values of pairs of quantitative features;

developed an algorithm for assessing the informativeness of a set of features and decision making system based on a stability measure;

developed an algorithm for reducing the feature space based on generalized estimates to eliminate cases of uncertainty in the data.

Implementation of the research results. Based on the scientific results obtained by the developed algorithms and their software, for decision making in the presence of uncertainty in the data:

an algorithm for calculating intervals of allowable values for pairs of quantitative features, an algorithm for assessing the reliability of decisions made in the presence of uncertainty in the data was used at the Scientific and Practical Center for Intelligent Software Systems at the National University of Uzbekistan named after Mirzo Ulugbek, (certificate of the Ministry of Digital Technologies of the Republic of Uzbekistan dated May 16, 2024 No. 33-8/3190). As a result, in the process of creating a database on pulmonary tuberculosis, the time for data verification and decision-making was reduced by 10-15%;

decision making system created on the basis of an algorithm for reducing the feature space based on generalized estimates to eliminate cases of uncertainty in the data and a parallel algorithm for the process of forming latent features was used in the State Unitary Enterprise “Center for Scientific, Technical and Marketing Research UNICON.UZ” (certificate of the Ministry of Digital Technologies of the Republic of Uzbekistan dated May 16, 2024 No. 33-8/3190). As a result, the time for data processing and decision-making was reduced by 15-20%.

Publication of research results. A total of 35 scientific papers were published on the research topic, of which 9 articles were published in recommended scientific editions for publication of basic scientific results of dissertations by Supreme Attestation Commission of the Republic of Uzbekistan, including 4 in national and 5 in foreign journals and 5 certificates for computer software.

Structure and volume of the dissertation. The dissertation consists of an introduction, four chapters, conclusion, the list of used literature and appendix. The volume of dissertation is 106 pages.

ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙХАТИ
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I бўлим (I часть; I part)

1. Lolaev M., Madrakhimov Sh., Makharov K., Saidov D. Choosing allowability boundaries for describing objects in subject areas. IAES International Journal of Artificial Intelligence, vol. 13, No. 1, March 2024, pp. 329-336. (№3, Scopus, IF=3.9).
2. Madrakhimov Sh., Makharov K. Informativeness of Feature Sets in Data with Missing Values. In: Aliev, R.A., et al. 12th World Conference “Intelligent System for Industrial Automation”. Lecture Notes in Networks and Systems, vol 718. 2024, pp 30-39. Springer, Cham. (№3, Scopus, IF=0.9).
3. Мадрахимов Ш. Ф., Махаров К. Т. Классификация объектов выборки с пропусками в данных // Проблемы вычислительной и прикладной математики. – 2023. – № 1(46). – С. 78-87. (05.00.00; №23)
4. Махаров К. Т. Построение базы нечётких правил в интеллектуальных системах для поддержки принятия решений. Colloquium-journal, no. 11 (170), 2023, pp. 13-16. (№12; Index Copernicus; IF=86.80).
5. Мадрахимов Ш. Ф., Махаров К. Т., Хуррамов А. Х. Отбор информативных признаков с учётом пропусков в данных // Проблемы вычислительной и прикладной математики. – 2022. – № 4(42). – С. 147-157. (05.00.00; №23)
6. Navruzov E.R., Makharov Q.T., Raximova M.A. Sinflarga ajratish masalasida alomatlarning informativligini baholash va ularni tanlash. Fan va jamiyat, 2022 (1), 27-30 b. (05.00.00; №37)
7. Shavkat Madrakhimov, Kodirbek Makharov, Musulmon Lolaev. Data preprocessing on input. AIP Conference Proceedings 2365, 030003, 2021. (№3, Scopus; IF=0.5)
8. Madrakhimov Sh.F., Rozikhodjaeva G.A. and Makharov K.T. “Construction of fuzzy inference rules for medicine diagnostics problems”, Journal of Physics: Conf. Ser. 2032 012032, 2021. (№3, Scopus; IF=1.2)
9. Махаров К.Т., Розыходжаева Г.А. Изучение процессов преждевременного старения сердечно-сосудистой системы методами машинного обучения. Вестник Ташкентской Медицинской Академии. №11, 2023, 184-186 стр. (14.00.00; №13)

II бўлим (II часть; II part)

10. Махаров К.Т., Хуррамов А.Х. Синф объектлари устунлик қилувчи интервалларни қуриш асосида берилганлардаги яширин қонуниятларни аниқлаш. Статистика ва унинг тадбиқи республика илмий-амалий семинари материаллари. Тошкент, ЎзМУ, 2013, 270-273 б.

11. Makharov K., Mirzaev J. Approximation of target parameters in problems of forecasting saline lands. Амалий математика ва инфорлатион технологияларнинг долзарб муаммолари – Ал-Хоразмий 2018. Тошкент-2018. 47-48 б.
12. Madrakhimov Sh.F., Makharov K.T. Analysis of the aging rate using Data Mining methods. International conference frontier in mathematics and computer science. Tashkent, October 12-15, 2020, p. 64.
13. Madrakhimov Sh., Rozikhodjaeva G., Makharov K. The use of Data Mining methods for estimating vascular aging. EAS Congress 2020. Virtual Congress, October 04-07, 2020.
14. Raximova M., Maxarov Q. Berilganlarda tushurib qoldirilgan qiymatlar bo'lganda informativ alomatlarini tanlash. "Matematik modellashtirish, hisoblash matematikasi va dasturiy ta'minot injeneriyasining dolzarb muammolari" respublika ilmiy-amaliy anjumani. Qarshi, 23-24 oktabr, 2020, 215-217 b.
15. Madraximov Sh., Maxarov Q. Qarish suratini tahlil qilishda data mining usullaridan foydalanish. "Matematik modellashtirish, hisoblash matematikasi va dasturiy ta'minot injeneriyasining dolzarb muammolari" respublika ilmiy-amaliy anjumani. Qarshi, 23-24 oktabr, 2020, 191-193 b.
16. Madrakhimov Sh.F., Makharov K.T. Building a base of fuzzy rules of an intelligent system for estimating person's successful aging. The 2020 International Conference on Computational Intelligence, Information Technology and Systems Research (CITS 20). Lublin, Poland, December 17-20, 2020, p. 21-22.
17. Махаров К.Т., Розыходжаева Г.А., Мадрахимов Ш.Ф. Методы интеллектуального анализа данных в изучении маркеров старения. Кардиологический вестник, спецвыпуск 2021, 164-стр.
18. Madraximov Sh., Maxarov Q. Intellectual tahlil usullari bilan insonning qarish suratini aniqlash masalalarida alomatlar fazosini qisqartirish. "Matematik fizika va matematik modellashtirishning zamonaviy muammolari" xalqaro ilmiy-amaliy anjumani. Qarshi, 3-4 dekabr, 2021, 12-13 b.
19. К.Т.Махаров, Г.А.Розыходжаева, Ш.Ф.Мадрахимов, З.Т.Икрамова. Выбор информативных признаков для классификации атеросклеротической нагруженности сонных артерий. Лучевая диагностика и терапия. №S(13), 2022 год, 189-190 стр.
20. Madrakhimov Sh.F., Makharov K.T. Determination of a set of informative features in the problem of estimating person's successful aging. Conference of young scientists "Mathematics, mechanics and intellectual technologies Tashkent-2022", April 21-22, 2022, p. 29-30.
21. Н.А.Игнатъев, Ш.Ф.Мадрахимов, К.Т.Махаров, А.Х.Тойчиев. Построение интеллектуальной системы диагностики COVID-19. Инфекция, иммунитет и фармакология. №3, 2022. Спец.выпуск. Материалы международной научно-практической конференции

- "Биология, этиология и физиология коронавируса COVID-19". Ташкент, 20 май, 2022, 319-320 стр.
22. Madrakhimov Sh.F., Makharov K.T. Search for hidden patterns in the diagnosis of aging index. International conference "Mathematical analysis and its applications in modern mathematical physics". Samarkand, September 23-24, 2022, p. 170-172.
 23. Madrakhimov Sh., Mukhamedova D., Saidov D., Makharov K. Application of artificial intelligence methods in archeological artifacts. International Conference 3D Digital Silk Road Project (3DSR'2022), Tashkent, Chirchik, Samarkand, Uzbekistan, September 7-9, 2022.
 24. Shavkat Madrakhimov, Kodirbek Makharov, Doniyor Saidov. Feature sets informativeness with data gaps. International Conference 3D Digital Silk Road Project (3DSR'2022), Tashkent, Chirchik, Samarkand, Uzbekistan, September 7-9, 2022.
 25. Мадрахимов Ш.Ф., Махаров К.Т. Классификация объектов выборки с пропусками в данных. Современное состояние и перспективы развития цифровых технологий и искусственного интеллекта. Сборник докладов республиканской научно-технической конференции, Самарканд, 26-27 октября 2022 г., 61-69 стр.
 26. Махаров Қ.Т. Tasniflash qobiliyatini oshirish uchun alomatlarni o'zgartirish. "Matematik modellashtirish va axborot texnologiyalarining dolzarb masalalari" xalqaro ilmiy-amaliy anjuman tezislari to'plami, Nukus, 2-3 may, 2023, 57-59 b.
 27. Madraximov Sh.F., Maxarov Q.T. Parallelizm va uni mashinali o'rgatish masalalarida qo'llash. Matematik modellashtirish, algoritmlash va dasturlashning dolzarb muammolari. Respublika ilmiy texnik anjumani, Toshkent, 14-15-sentabr, 2023 yil. 41-44 b.
 28. Rakhimov B.B., Alimov A.A., Rakhimova M.A., Makharov K.T. Review of machine learning methods for predicting the severity of covid-19. 8th international conference "Actual problems of applied mathematics and information technologies" – Al-Khwarizmi 2023, Samarkand, September 25-26, 2023, p. 286.
 29. Madrakhimov Sh.F., Makharov K.T. Parallel algorithms for the building of latent feature space. 8th international conference "Actual problems of applied mathematics and information technologies" – Al-Khwarizmi 2023, Samarkand, September 25-26, 2023, p. 280.
 30. Khasanova K., Rakhimova M., Makharov K., Mahkamova N., Ibodullaev Z. Ischemic heart disease prediction using machine learning methods. 8th international conference "Actual problems of applied mathematics and information technologies" – Al-Khwarizmi 2023, Samarkand, September 25-26, 2023, p. 278.
 31. Maxarov Q.T., Mirzayev A.I., Raximova M.A., Navruzov E.R., Tuliyeu U.Yu. Obyektlarning umumlashgan bahosini hisoblash dasturi. // O'zbekiston Respublikasi adliya vazirligi huzuridagi intellektual mulk agentligi. Elektron

- hisoblash mashinalari uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnoma № DGU 09800. 25.11.2020.
32. Raximov B.B., Maxarov Q.T., Alimov A.A., Raximova M.A. COVID-19 virusi bilan kasallangan bemorlarning kasallik og'irlik darajasini aniqlash dasturi. // O'zbekiston Respublikasi adliya vazirligi huzuridagi intellektual mulk agentligi. Elektron hisoblash mashinalari uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnoma № DGU 19106. 07.10.2022.
 33. Suyunov N.D., Rajabova N.X., Maxarov Q.T. Intellektual tahlil usullari asosida onkologik kasalliklarni davolashda qo'llaniladigan dori vositalari samaradorligini aniqlash dasturi (o'pka saratoni misolida). // O'zbekiston Respublikasi adliya vazirligi huzuridagi intellektual mulk agentligi. Elektron hisoblash mashinalari uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnoma № DGU 22865. 19.02.2023.
 34. Maxarov Q.T. To'liq bo'lmagan berilganlarda tasniflash dasturi. // O'zbekiston Respublikasi adliya vazirligi huzuridagi intellektual mulk agentligi. Elektron hisoblash mashinalari uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnoma № DGU 25129. 26.04.2023.
 35. Raximov B.B., Maxarov Q.T., Raximova M.A., Alimov A.A. To'liq bo'lmagan berilganlarda koronavirus kasalligini aniqlash dasturi. // O'zbekiston Respublikasi adliya vazirligi huzuridagi intellektual mulk agentligi. Elektron hisoblash mashinalari uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnoma № DGU 26813. 14.07.2023.

Avtoreferat “Hisoblash va amaliy matematika muammolari” ilmiy jurnali tahririyatida tahrirdan o’kazildi va o’zbek, rus va ingliz tillaridagi matnlarini mosligi tekshirildi.

