

## **РЕЗЮМЕТА И ПРИНОСИ НА НАУЧНИТЕ ТРУДОВЕ**

*на*

*доц. д-р инж. Десислава Антонова Иванова*

за участие в конкурс за заемане на академична длъжност „Професор“  
в област на висше образование: 4. „Природни науки, математика и информатика“  
Професионално направление: 4.6. Информатика и компютърни науки  
Специалност „Информатика“  
обявен в ДВ брой 2 / 07 януари 2025 г.

**Хабилитационен труд – резюмета на научни публикации в издания, които са реферирани и индексирани в световноизвестни бази данни с научна информация (Web of Science и Scopus)**

### ***ПОДХОДИ, МЕТОДИ И АЛГОРИТМИ ЗА АНАЛИЗ НА ГОЛЕМИ БИОМЕДИЦИНСКИ ДАННИ В ПОДКРЕПА НА ПРЕЦИЗНАТА МЕДИЦИНА С ИЗПОЛЗВАНЕ НА ИЗКУСТВЕН ИНТЕЛЕКТ***

През последните години водещи учени, изследователи и анализатори посочват големите данни (Big Data) като революция в научните изследвания и една от най-перспективните тенденции в областта на ИТ. Това даде тласък на интензивното развитие на методите и технологиите за обработка на големи масиви от данни в последните години и доведе до радикални промени в парадигмите за научни изследвания.

В глобален мащаб, интензивният прогрес в ИКТ доведе до натрупването и съхранението на огромни масиви данни, в общия случай с ниска информационна плътност, които се превръщат в основен източник на знания. Новата парадигма за научни изследвания „Data-Intensive-Scientific-Discovery“ (DISD) радикално променя начина за провеждане на експериментите и откриване на знания. Вместо да се планира експеримента и след това да се анализират данните, при новата парадигма натрупаните огромни масиви данни се подлагат на анализ, при който се търсят значими корелации и се прилагат иновативни методи за откриване на нови знания. Прилагането на този радикално нов подход за провеждане на научни изследвания е изключително

Резюмета на научните трудове и приносите на доц. д-р инж. Десислава Антонова Иванова

полезен за анализ на големи биомедицински данни в подкрепа на прецизната медицина, с висока добавена стойност за обществото и нашето здраве.

Прецизната медицина е една от най-горещите теми в днешно време и включва диагностика и лечение на заболявания, които отчитат индивидуалния генетичен профил, спецификата на околната среда и начина на живот на индивида. Откриването на знания посредством анализ на данни (KDD) е от решаващо значение за идентифицирането на сложни ДНК аномалии, свързани с генетични заболявания и рак. Капацитетът на ДНК базите данни нараства експоненциално през последното десетилетие. Прецизната медицина започва с геномика и разчита на платформите „omics“ за анализ и интерпретация на многомащабни данни, както и на новите методи и средства за анализ на големи данни с използване на изкуствен интелект.

- I. **D. Ivanova**, *Proactive Intelligent System for Early Detection of Thyroid Anomalies*, AIP Conference Proceedings 2505, 020002 (2022), ISBN 978-073544396-9, <https://doi.org/10.1063/5.0100631>; [Scopus - Document details - Proactive Intelligent System for Early Detection of Thyroid Anomalies](#)

Бързото развитие на компютърните и софтуерните науки засяга всички сфери на живота, включително областта на медицината и здравеопазването. Ранната диагностика на заболяванията и персонализираната терапия изискват използването на съвременни информационни технологии и изкуствен интелект, които да подкрепят лекарите при вземането на решения.

Статията предлага системна архитектура и решение на проактивна интелигентна система, която е предназначена за ранна диагностика на заболявания на щитовидната жлеза. Системата се основава на методи и алгоритми с изкуствен интелект за анализ на големи биомедицински данни за ранното откриване на аномалии на щитовидната жлеза и персонализирана терапия.

Проактивната система предоставя възможност за 1) Оценка на състоянието и риска от заболяване на щитовидната жлеза, чрез разработения скоринг модул (scoring module), който се състои от *модел на Зулевски и модел за оценка на състоянието на щитовидната жлеза след прекаран COVID-19*. 2) Прогнозна диагноза за състоянието на щитовидната жлеза от модула за функционални изменения (functional disorders module). В този модул на

системата е разработен алгоритъм за анализ на данни от кръвни изследвания, даващ прогноза за функционалното състояние на щитовидната жлеза.

3) Прогнозна диагноза за състоянието на щитовидната жлеза от модула за структурни изменения (structural disorders module). Този модул на системата, използва алгоритъм с изкуствен интелект, с помощта на който е разработен класификатор за откриване на структурни аномалии на щитовидната жлеза от анализ на големи биомедицински данни от ехографски изследвания.

Трите нива на проактивната интелигентна система предоставят възможност за прогнозиране на функционални и структурни аномалии на щитовидната жлеза, заедно с прогнозиране на риска от заболяване на щитовидната жлеза.

В България процентът на пациентите с аномалии на щитовидната жлеза са милион и половина, заедно с факта, че броят на медицинските специалисти намалява, са причина да се проправи път към проектиране и развитие на предложената проактивна интелигентна система. Процесът на разработване на предложената проактивна интелигентна системата изисква интердисциплинарна работа, съчетаваща ИТ експертиза и медицински компетенции.

## ПРИНОСИ:

1. Извършен е анализ на научната продукция, свързана със съвременните постижения в областите на информатиката, големите данни и прецизната медицина чрез прилагане на библиометричен подход и систематичен литературен обзор.
2. Разработени са функционалните изисквания, алгоритмите и моделите с изкуствен интелект на проактивна система за ранна диагностика и откриване на аномалии на щитовидната жлеза.
3. Предложена е системна архитектура на проактивна интелигентна система за ранна диагностика на аномалии на щитовидната жлеза.

- II. **D. Ivanova**, *Artificial Intelligence in Internet of Medical Imaging Things: The Power of Thyroid Cancer Detection*, IEEE, 28 October 2018, International Conference on Information Technologies (InfoTech), 2018, <https://ieeexplore.ieee.org/document/8510725>, DOI:[10.1109/InfoTech.2018](https://doi.org/10.1109/InfoTech.2018).

[8510725; Scopus - Document details - Artificial intelligence in internet of medical imaging things: The power of thyroid cancer detection](#)

Развитието на информатиката и компютърната наука е в основата на глобалната промяна на информационната среда, здравеопазването и индустриалното развитие в световен мащаб. Интернет на нещата (IoT) размива границата между реалната и виртуалната среда. Това е не само инструмент за разработване на новия дигитален свят, но и инструмент за развитие на бъдещо общество на споделено знание посредством милиони свързани устройства. Тези нови устройства не само ни помагат да разберем повече за нашето здраве, но и събират големи данни от самите нас.

Интернет на нещата в медицината (IoMT) и големите данни (Big Data) са две области, които могат да революционизират изследванията в медицината. Ядрото на тези тенденции е изкуственият интелект (AI), който добавя стойност към медицината.

Статията предлага подход за откриване на рак на щитовидната жлеза, базиран на изкуствен интелект в екосистемата Интернет на нещата в медицинската образна диагностика (IoMIT). Ултразвуковите изображения, събрани в екосистемата на IoMIT, е най-добрият начин за диагностика на рак на щитовидната жлеза. Сегментирането на изображението и откриването на доброкачествени и злокачествени тиреоидни възли е важна част от предложения подход. Той е реализиран в среда на Apache Spark с помощта на библиотеката за машинно обучение MLlib, базиран на конволюционни невронни мрежи (CNN).

### **ПРИНОСИ:**

1. Извършен е анализ на научната продукция, свързана със съвременните постижения в областите на Интернет на нещата в медицината и образната диагностика чрез прилагане на библиометричен подход и систематичен литературен обзор.
2. Предложен е подход за откриване на рак на щитовидната жлеза основаващ се на анализ и откриване на аномалии от ултразвукови медицински изображения в екосистемата Интернет на нещата.

3. Предложеният подход е верифициран и валидиран на основата на програмна реализация и симулационни експерименти в среда на Apache Spark с помощта на библиотеката за машинно обучение MLlib и използване на конволюционни невронни мрежи (CNN).

III. **D. Ivanova**, P. Borovska, S. Zahov, *Development of PaaS using AWS and Terraform for medical imaging analytics*, AIP Conference Proceedings 2048, 060018 (2018); <https://doi.org/10.1063/1.5082133>; [Scopus - Document details - Development of PaaS using AWS and Terraform for medical imaging analytics](#); [Development of PaaS using AWS and Terraform for medical imaging analytics-Web of Science Core Collection](#)

Статията представя концепцията и съвременния напредък в изчислителните архитектури, моделите и услугите в облака. Целта е да се разработи облачна архитектура и да се представи, как една абсолютно сигурна, бърза и мощна облачна изчислителна среда, може да бъде изградена по лесен и икономичен начин, за да служи като краен продукт. За тази цел е създадена облачна среда чрез използване на *нов, модерен и развиващ се подход*, инфраструктура като код (IaC) чрез написване на Terraform код със съответния облачен източник, AWS “Amazon Web Services”. Разработената облачна инфраструктура ще осигури платформа като услуга „PaaS (Platform as a Service)“ за анализ на медицински изображения и ще поддържа всички необходими софтуерни инструменти и пакети.

### **ПРИНОСИ:**

1. Извършен е анализ на научната продукция, свързана със съвременните постижения в областите на облачните технологии и услуги чрез прилагане на библиометричен подход и систематичен литературен обзор.
2. Предложен е нов и модерен подход за изграждане на изчислителни облачни среди с използване на инфраструктура като код (IaC), Terraform код със съответния облачен източник, AWS “Amazon Web Services”.

3. Проектирана е облачна архитектура, базираща се на предложния нов подход.
4. Разработена е платформа като услуга „PaaS (Platform as a Service)“ за анализ на медицински изображения.

IV. **D. Ivanova, V. Zlatanov, *Anomaly detection of thyroid hormones disorder based on machine learning*, AIP Conference Proceedings 2505, 020003 (2022); <https://doi.org/10.1063/5.0100632>; [Scopus - Document details - Anomaly Detection of Thyroid Hormones Disorder Based on Machine Learning](#)**

Статията предлага подход за откриване на аномалии при нарушения на тиреоидните хормони и аномалии в медицинските изображения на щитовидната жлеза въз основа на машинно обучение. Предложеният подход е комбинация от две различни стъпки – анализ на кръвен тест и класификация на ултразвуково изображение.

Първата стъпка от подхода е анализът на кръвта за хормонални нарушения. Състои се от две фази – фаза на предварителна обработка и фаза на машинно обучение. Отговорностите на фазата на предварителна обработка включват филтриране, почистване и разделяне на данни, както и избор на характеристики. През фазата на машинно обучение се осъществяват процеси на обучение, класификация и прогнозиране. Целта на този модул е да осигури ранен етап на наблюдение на пациент с помощта на кръвен тест. Резултатите от теста се качват ръчно в системата и след това се класифицират в една от трите изходни категории – нормални, хипертиреоидизъм и хипотиреоидизъм. В случая ядрото на тази фаза е алгоритъмът на Naive Bayes.

Втората стъпка от подхода е анализ на ултразвуковото изображение на щитовидната жлеза за откриване на аномалии. Подобно на анализа на кръвния тест, се прилага подход, който включва предварителна обработка и фаза на машинно обучение. В етапа на предварителна обработка наборът данни от ултразвукови изображения се качва в системата и пикселните данни от всяко изображение се извличат. След това следва фазата на тесни места, в която стойностите на пикселите на изображенията се използват като вход за

замразените слоеве на модела. Изходните стойности на този процес се използват като вход за дълбоката невронна мрежа в следващата стъпка.

Етапът на машинно обучение е разделен на две подкатегории. Първият е обучението на данните, при което изображенията се разпределят в пет предварително дефинирани категории на EU-TIRADS. Това се прави с помощта на остатъчната дълбока невронна мрежа ResNet-50, интегрирана в рамката ML.NET. Наборът от данни е разделен на 80% набор за обучение и 20% тестов набор. Наборът за обучение се използва за машинно обучение на модела, а наборът за тестване оценява точността на обученния модел. Процесът се повтаря до постигане на добър резултат. В края на тази подкатегория се генерира персонализиран модел TensorFlow. Другата подкатегория е фазата на прогнозиране. Неговата цел е да класифицира ново ултразвуково изображение на щитовидната жлеза (въведено от потребителя) към една от петте категории на обученния модел. Тази цел се постига чрез изчисляване на вероятността за съвпадение между входното изображение и всяка от категориите EU-TIRADS. Накрая изображението се отнася към категорията с най-висока вероятност за съответствие.

### **ПРИНОСИ:**

1. Предложен е подход за откриване на аномалии при нарушения на тиреоидните хормони и аномалии в медицинските изображения на щитовидната жлеза въз основа на машинно обучение.
2. Разработен е класификатор за откриване на структурни аномалии на щитовидната жлеза по системата EU-TIRADS с данни от образна диагностика и използване на изкуствен интелект.
3. Предложеният подход и класификатор са верифицирани и валидирани на основата на програмна реализация и симулационни експерименти в среда на .NET 4.6.1, .NET Core 2.2 и ML.NET 1.5.1 и използване на дълбока невронна мрежа ResNet-50 и алгоритъма на Naive Bayes.

V. **D. Ivanova, B. Nenchovski, *Experimental Framework and Graphical User Interface for Research in Bioinformatics: GRAY – Gene Rapid Analysis***, DOI: 10.1109/InfoTech49733.2020.9210973, Conference: 2020 International Conference on Information Technologies (InfoTech),

<https://ieeexplore.ieee.org/document/9210973>; [Scopus - Document details - Experimental Framework and Graphical User Interface for Research in Bioinformatics: GRAY - Gene Rapid Analysis](#)

Фокусът на тази статия е върху численото представяне на данни за гени в областта на биоинформатиката. Представени са ползите от използването на дълбоко обучение, които са свързани с факта, че този тип алгоритми допълват ДНК анализа, тъй като те отчитат индивидуалната променливост в гените и могат да помогнат при лечението и превенцията на заболявания чрез определяне на генетичната предразположеност към заболявания.

Статията показва процесът на създаване на лесен и удобен графичен потребител интерфейс (GUI), който осигурява достъп до публични биологични бази данни, позволявайки на потребителя да редактира данните и да ги анализира с алгоритми за машинно обучение без изискващи познания по програмиране или специализирани хардуер.

Модерни софтуерни библиотеки и най-добрите практиките за програмиране са използвани за проектиране на архитектурата и създаването на графичния интерфейс.

Така предложената експериментална рамка и графичен интерфейс е насочен към изследователите, които работят с генни последователности и техния анализ. Разработката представена в статията предлага софтуерен продукт с отворен код “GRAY - Gene Rapid AnalYsis”.

Тази статия представя експериментална рамка, базирана на изкуствени невронни мрежи като многослойния перцептрон, с основана цел класификация и анализ на гени. Тя е насочена към намаляване на времето, необходимо за обучение на такава мрежа, в допълнение към понижаване на хардуерните изисквания за нейната виртуализация, без да се жертва точността.

### **ПРИНОСИ:**

1. Предложена е работна рамка за анализ на геномни данни с изкуствен интелект.
2. Създаден е лесен и удобен графичен интерфейс, който осигурява достъп до публични биологични бази данни и последващ анализ с алгоритми за



машинно обучение, без изискващи познания по програмиране или специализирани хардуер.

3. Разработен е софтуер с отворен код “GRAY - Gene Rapid AnaLYsis” за анализ на големи геномни данни.
4. Работоспособността на предложения софтуерен продукт „GRAY“ е верифицирана и валидирани на базата на експерименти с различни геномни данни.

VI. **D. Ivanova, J. Staeva, *Structural Similarity Analysis of SARS-CoV-2 and Human Proteins***, AIP Conference Proceedings 2505, 020001 (2022); <https://doi.org/10.1063/5.0100629>; [Scopus - Document details - Structural Similarity Analysis of SARS-CoV-2 and Human Proteins](#)

Целта на тази статия е да предостави анализ на структурното сходство на SARS-CoV-2 и човешки протеини. За тази цел е предложена експериментална рамка, базирана на множество иновативни платформи за биоинформатика.

Този експеримент има за цел да определи кои човешки и SARS-2 протеини имат най-висок процент на структурно сходство и да измери този процент. За да изпълним тази задача се предлага доста нетрадиционен работен процес в четири фази: 1) Избор на подходяща база данни, която съдържа човешки и SARS-2 протеини. 2) Филогенен анализ, за да визуализират протеиновите двойки с най-ниско еволюционно разстояние. 3) Подравняване на последователностите на най-близките двойки човешки и SARS-2 протеини, като се използват резултатите от филогенетичното дърво, за да се измери техният процент на структурно сходство. 4) Визуализация и анализ на резултатите.

Резултатите от изследването се анализират по отношение на процента на идентичност между човешки и SARS-2 протеинови двойки. Обобщението на резултатите показва, че процентът на идентичност между човешки и SARS-2 протеинови двойки варира в диапазона от 7,6 % до 23,6 %. Степента на сходство варира между 8,5 % и 35,4 %. Протеиновата двойка с най-висок резултат за идентичност и сходство се състои от протеини с по-къси последователности.

### **ПРИНОСИ:**

1. Предложена е работна рамка за анализ на структурното сходство на SARS-CoV-2 и човешки протеини.
2. Създаден е работен процес покриващ всички фази на конвейера за анализ на структурното сходство на SARS-CoV-2 и човешки протеини.
3. Работоспособността на предложената работна рамка за анализ на структурното сходство на SARS-CoV-2 и човешки протеини е верифицирана и валидирана на базата на множество проведени експерименти с различни геномни данни.

- VII. **D. Ivanova, V. Kadurin**, A new proposed software development methodology for healthcare industry, AIP Conference Proceedings 2333, 030010 (2021); <https://doi.org/10.1063/5.0042261>; [Scopus - Document details - A new proposed software development methodology for healthcare industry](#); [A New Proposed Software Development Methodology for Healthcare Industry-Web of Science Core Collection](#)

Дигитализирането на здравните системи и процеси е от съществено значение за повишаване на безопасността на пациентите и намаляване на медицинските грешки. Комплексната логика на здравеопазването и нулевата толерантност към грешки изискват специални процеси за създаване на медицински софтуер. Целта на тази статия е да подчертае и развие основните разлики между всички значими и широко използвани методологии за разработка на софтуер, тъй като всички те имат своите полезни свойства и недостатъци в различни ситуации. В статията е обсъдена общата използваемост на методологиите, включително описание на това, как се постига култура, непрекъснатата интеграция и доставка с всяка една от тях.

Успешното внедряване на здравен софтуер е трудна задача поради естеството на домейна. То има свойства и функции, които са много различни от всеки друг софтуерен продукт. През последните две десетилетия, приложенията за здравеопазване и използването на информационни системи в здравната индустрия нараснаха бързо. Основната цел на тези приложения е подобряването на качеството на медицинската помощ и услуга.

В тази статия е предложена нова методология за разработка на софтуер за здравеопазването. Тя включва 4 отделни фази, които са изпълняват се последователно, но имат вътрешни итеративни модели – 1) план за анализ, 2) архитектура, 3) разработка, 4) валидиране и проверка. Това носи предимствата както на водопадния, така и на гъвкавия подход. Този комбиниран подход отговаря специално на нуждите на сложната сфера на здравеопазването.

### **ПРИНОСИ:**

1. Извършен е анализ на научната продукция, свързана със съвременните постижения в областта на софтуерното инженерство и методологиите за разработка на софтуер чрез прилагане на библиометричен подход и систематичен литературен обзор.
2. Предложена е нова методология за разработка на софтуер в здравеопазването, включваща 4 отделни фази, които се изпълняват последователно, но имат и вътрешни итеративни модели - план за анализ, архитектура, разработка, валидиране и проверка.

**ПРИНОСИ от научна публикация в издания, които са реферирани и индексирани в световноизвестни бази данни с научна информация (Web of Science и Scopus), извън хабилитационния труд за група от**  
**ПОКАЗАТЕЛИ Г**

Статиите от комплекта документи *Г1, Г2, Г3, Г4, Г5* са свързани със съвременните постижения в използването на цифрови технологии за интерактивно представяне на българското културноисторическо наследство и по-конкретно: *UI/UX дизайн, програмирането за мобилни устройства и създаването на атрактивни мобилни приложения в света на културно-историческото наследство.*

### **ПРИНОСИ**

1. Предложен е концептуален модел „Виртуален площад на българското културно-историческо наследство“.

2. Създаден е сценарият за виртуални обиколки във виртуалния площад с автобус „*Hop on, hop off*“ за интерактивно представяне на българското културно-историческо наследство с използване на модерни дигитални технологии с виртуална и добавена реалност.
3. Предложен е методологичен процес на разработка на мобилни приложения, отразявайки предизвикателствата и най-добрите практики в областта.
4. Предложена е архитектура на мобилно приложение за интерактивно представяне на българското културно историческо наследство.
5. Разработено е мобилно приложение за българското културно историческо наследство с игровизация и геопозициониране, което предлага на потребителите интерактивен и удобен начин да откриват и опознават богатата история и култура на България.
6. Извършен е анализ на научната продукция, свързана със съвременните постижения в областта на UI/UX дизайн чрез прилагане на библиометричен подход и систематичен литературен обзор.
7. Предложена е стратегия за използваемост и насоки за развитие на уеб портал за българското културно и историческо наследство.
8. Предложен е хибридният подход, който съчетава доказани и работещи методологии и иновативни процеси, фокусирани върху дизайн, ориентиран към потребителя.
9. Предложена е иновативна рамка, ориентирана към потребителя, за разработка на софтуер за интерактивно представяне на българското културно-историческо наследство.

Представените статии *Г6* и *Г7* са свързани с изследвания в областта на екосистемата Интернет на нещата, анализа на големи данни и прилагането на методите на изкуствения интелект. Приносите са както следва:

## **ПРИНОСИ**

1. Предложена е рамка и архитектура на интелигентна система за оценка и мониторинг на качеството на въздуха в екосистемата IoT.
2. Създаден е прогнозен модел с машинно обучение, базиран на многослоен перцептрон за оценка качеството на въздуха.

3. Разработен е графичен интерфейс на интелигентната система, които да предоставя възможност на потребителите да наблюдават прогнозираните стойности за качеството на въздуха.
4. Предложена е хибридна концептуална рамка на взаимодействието между системите за контрол на управлението и анализа на големи данни, която позволява успешно постигане на стратегическите цели на компанията и оценка на влиянието върху четирите лоста за контрол.

**ПРИНОСИ от научна публикация в издание с импакт фактор (IF на Web of Science) и/или с импакт ранг (SJR на Scopus), за група от ПОКАЗАТЕЛИ 3**

Статиите от комплекта документи 31, 32 и 33 са свързани със сигурността на екосистемата IoT, най-новите тенденции в ИТ и най-новите постижения на науката в подкрепа на персонализираната медицина за оценка на риска от рак на гърдата и анализ на големи геномни данни. Приносите са както следва:

**ПРИНОСИ**

1. Предложен е концептуалният модел на платформа за интелигентни решения за сигурност в екосистемата IoT.
2. Създадена е концептуална софтуерна архитектура на интелигентна система за откриване на атаки изградена от две секции: 1) за машинно обучение (ML) и 2) оперативна секция.
3. Предложен е подход за изграждане на паралелни диференцирани диагностично аналитични работни потоци за казуса на откриване на прониквания и атаки.
4. Направена е верификация и валидация на предложената интелигентна система на базата на множество експерименти за казуса на диагностиката.
5. Предложена е класификацията на рисковите фактори за рак на гърдата и наличните модели и инструменти за оценка на риска.
6. Разработени са модели за оценка на риска от рак на гърдата. Създаден е биоинформатичен работен процес за откриване на знания от данни (KDD) за оценка на риска от рак на гърдата и определяне на типа риск: висок, повишен и нисък риск.

7. Направена е верификация и валидация на разработените модели и предложени биоинформатичен работен процес за оценка на риска от рак на гърдата на базата на експериментални изследвания. Получените резултати с точност до 93% при оценката от риска от рак на гърдата доказват конкурентността на предложените решения.
8. Предложена е интелигентен метод за адаптивно *in silico* откриване на знания, базиран на анализ на големи геномни данни.
9. Предложена е архитектура на софтуерната система, изградена на базата на метода.
10. Направена е верификация и валидация на метода за казуси на работни процеси за диференциран описателен анализ за откриване и картографиране на прокариотни и еукариотни гени и работни процеси за диференциран диагностичен анализ за откриване на генна мутация, свързана с рак на гърдата.

## **ABSTRACTS AND CONTRIBUTIONS OF SCIENTIFIC PAPERS**

on

*Associate Professor Desislava Antonova Ivanova, Ph.D., Eng*

for participation in a competition for the academic position "Professor"  
in the field of higher education: 4. "Natural sciences, mathematics and informatics"

Professional direction: 4.6. Informatics and Computer Science

Specialty "Informatics"

announced in SG No. 2 / January 7, 2025

**HABILITATION THESIS - summaries of scientific publications which are referenced and indexed in world-famous databases with scientific information (Web of Science and Scopus)**

### ***APPROACHES, METHODS AND ALGORITHMS FOR ANALYSIS OF BIG BIOMEDICAL DATA IN SUPPORT OF PRECISION MEDICINE USING ARTIFICIAL INTELLIGENCE***

In recent years, leading scientists, researchers, and analysts point to Big Data as a revolution in scientific research and one of the most promising trends in IT. This gave impetus to the intensive development of methods and technologies for processing big data sets in recent years and led to radical changes in scientific research paradigms.

On a global scale, intensive progress in ICT has led to the accumulation and storage of big data sets, generally with low information density, which are becoming the main source of knowledge. The new research paradigm "Data-Intensive-Scientific-Discovery" (DISD) is radically changing the way we conduct experiments and discover knowledge. Instead of planning the experiment and then analyzing the data, in the new paradigm, the accumulated big arrays of data are subjected to analysis that looks for significant correlations and applies innovative methods to discover new knowledge. Applying this radical new approach to conducting scientific research is extremely useful for the analysis of big biomedical data in support of precision medicine, with high added value for society and our health.

Precision medicine is one of the hottest topics today and involves the diagnosis and treatment of diseases that consider the individual's genetic profile, the environment specifics, and the individual lifestyle. Knowledge data discovery

(KDD) is critical to the identification of complex DNA abnormalities associated with genetic diseases and cancer. The capacity of DNA databases has grown exponentially over the past decade. Precision medicine starts with genomics and relies on 'omics' platforms to analyse and interpret big data, as well as new methods and tools for big data analytics using artificial intelligence.

- I. **D. Ivanova**, *Proactive Intelligent System for Early Detection of Thyroid Anomalies*, AIP Conference Proceedings 2505, 020002 (2022), ISBN 978-073544396-9, <https://doi.org/10.1063/5.0100631>; [Scopus - Document details - Proactive Intelligent System for Early Detection of Thyroid Anomalies](#)

The rapid development of computer and software sciences affects all spheres of life, including the field of medicine and healthcare. Early diagnosis of diseases and personalized therapy require the use of modern information technologies and artificial intelligence to support doctors in decision-making.

The paper proposes a system architecture and solution of a proactive intelligent system that is designed for early diagnosis of thyroid diseases. The system is based on artificial intelligence methods and algorithms to analyse large biomedical data for the early detection of thyroid abnormalities and personalized therapy.

The proactive system provides an opportunity for 1) Risk evaluation of the thyroid disease, through the developed scoring module (scoring module), which consists of Zulevsky's model for evaluating the state of the thyroid gland after passing through COVID-19. 2) Predictive diagnosis of the state of the thyroid gland from the functional disorders module. In this module of the system, an algorithm has been developed for the data analysis from blood tests, giving a prediction of the functional state of the thyroid gland. 3) Predictive diagnosis of the state of the thyroid gland from the structural disorders module. This system module uses an artificial intelligence algorithm for detection of structural abnormalities of the thyroid gland from the analysis of big biomedical data of ultrasound examinations.

The three levels of the proactive intelligent system provide the ability to predict functional and structural abnormalities of the thyroid gland along with predicting the risk of thyroid disease.

In Bulgaria, the percentage of patients with thyroid abnormalities is one and a half million, together with the fact that the number of medical specialists is decreasing, these are reasons to pave the way for the design and development of the



proposed proactive intelligent system. The development process of the proposed proactive intelligent system requires interdisciplinary work combining IT expertise and medical competencies.

### CONTRIBUTIONS:

1. An analysis of the scientific output related to modern achievements in the fields of informatics, big data and precision medicine was carried out by applying a bibliometric approach and a systematic literature review.
2. The functional requirements, algorithms, and artificial intelligence models of a proactive system for early diagnosis and detection of thyroid gland abnormalities are developed.
3. A system architecture of a proactive intelligent system for early diagnosis of thyroid gland abnormalities is proposed.

- II. **D. Ivanova**, *Artificial Intelligence in Internet of Medical Imaging Things: The Power of Thyroid Cancer Detection*, IEEE, 28 October 2018, International Conference on Information Technologies (InfoTech), 2018, <https://ieeexplore.ieee.org/document/8510725>, DOI:[10.1109/InfoTech.2018.8510725](https://doi.org/10.1109/InfoTech.2018.8510725); [Scopus - Document details - Artificial intelligence in internet of medical imaging things: The power of thyroid cancer detection](#)

The development of informatics and computer science is at the heart of the global change of the information environment, health care and industrial development worldwide. The Internet of Things (IoT) is blurring the line between real and virtual environments. It is not only a tool to develop the new digital world, but also a tool to develop a future society of shared knowledge through millions of connected devices. These new devices not only help us understand more about our health, but also collect big data from ourselves.

The Internet of medical Things (IoMT) and Big Data are two areas that can revolutionize medical research. At the core of these trends is artificial intelligence (AI), which adds value to medicine.

The paper proposes an artificial intelligence-based approach for thyroid cancer detection in Internet of Medical Imaging Things (IoMIT) ecosystem. Ultrasound imaging collected in the IoMIT ecosystem is the best way to diagnose thyroid cancer. Image segmentation and detection of benign and malignant thyroid

modules is an important part of the proposed approach. It is implemented in an Apache Spark environment using the MLlib machine learning library based on Convolutional Neural Networks (CNN).

### **CONTRIBUTIONS:**

1. An analysis of the scientific production related to modern achievements in the fields of the Internet of medical Things and imaging diagnostics was carried out by applying a bibliometric approach and a systematic literature review.
2. An approach for thyroid cancer detection based on analysis and anomaly detection of ultrasound medical images in Internet of Things ecosystem is proposed.
3. The proposed approach is verified and validated based on program implementation and simulation experiments in an Apache Spark environment using the machine learning library MLlib and using convolutional neural networks (CNN).

III. **D. Ivanova**, P. Borovska, S. Zahov, *Development of PaaS using AWS and Terraform for medical imaging analytics*, AIP Conference Proceedings 2048, 060018 (2018); <https://doi.org/10.1063/1.5082133>; [Scopus - Document details - Development of PaaS using AWS and Terraform for medical imaging analytics](#); [Development of PaaS using AWS and Terraform for medical imaging analytics-Web of Science Core Collection](#)

The paper presents the concept and current advances in cloud computing architectures, models, and services. The goal is to develop a cloud architecture and to present how a secure, fast and powerful cloud computing environment can be built in an easy and economical way to serve as a final product. For this purpose, a cloud environment is created by using a new, modern, and evolving approach, Infrastructure as Code (IaC) by writing Terraform code with the corresponding cloud source, AWS "Amazon Web Services". The developed cloud infrastructure will provide a PaaS (Platform as a Service) for medical image analytics and will support all necessary software tools and packages.

### CONTRIBUTIONS:

1. An analysis of the scientific output related to modern achievements in the fields of cloud technologies and services was carried out by applying a bibliometric approach and a systematic literature review.
  2. A new and modern approach to building computing cloud environments using infrastructure as a code (IaC), Terraform code with the corresponding cloud source, AWS "Amazon Web Services" is proposed.
  3. A cloud architecture based on the proposed new approach is designed.
  4. A Platform as a service "PaaS" has been developed for the analysis of medical images.
- IV. **D. Ivanova, V. Zlatanov, *Anomaly detection of thyroid hormones disorder based on machine learning***, AIP Conference Proceedings 2505, 020003 (2022); <https://doi.org/10.1063/5.0100632>; [Scopus - Document details - Anomaly Detection of Thyroid Hormones Disorder Based on Machine Learning](#)

The paper proposes a machine learning-based approach to detect abnormalities in thyroid hormone disorders and abnormalities in the structure of thyroid gland. The proposed approach is a combination of two different steps – blood test analysis and ultrasound image classification.

The first step of the approach is the blood analysis for hormonal disorders. It consists of two phases – a pre-processing phase and a machine learning phase. Responsibilities of the pre-processing phase include data filtering, cleaning, and partitioning, as well as feature selection. During the machine learning phase, learning, classification, and prediction processes take place. The purpose of this module is to provide early-stage monitoring of a patient using a blood test. Test results are manually uploaded into the system and then classified into one of three output categories – normal, hyperthyroidism and hypothyroidism. In this case, the core of this phase is the Naive Bayes algorithm.

The second step of the approach is an ultrasound image analysis of the thyroid gland to detect abnormalities. Like blood test analysis, an approach is applied that includes pre-processing and a machine learning phase. In the pre-processing step, the dataset of ultrasound images is uploaded to the system and the pixel data of each image is extracted. Then follows the bottleneck phase, in which the pixel values of

the images are used as input to the frozen layers of the model. The output values of this process are used as input for the deep neural network in the next step.

The machine learning stage is divided into two subcategories. The first is data training, where images are classified into five predefined EU-TIRADS categories. This is done using the ResNet-50 residual deep neural network integrated into the ML.NET framework. The dataset is divided into 80% training set and 20% test set. The training set is used for machine training the model and the testing set evaluates the accuracy of the trained model. The process is repeated until a good result is achieved. At the end of this subcategory, a custom TensorFlow model is generated. The other subcategory is the prediction phase. Its goal is to classify a new thyroid ultrasound image (entered by the user) into one of the five categories of the trained model. This goal is achieved by calculating the probability of a match between the input image and each of the EU-TIRADS categories. Finally, the image is assigned to the category with the highest probability of matching.

#### **CONTRIBUTIONS:**

1. A machine learning-based approach for detecting abnormalities in thyroid hormone disorders is proposed.
2. A classifier was developed for the detection of structural abnormalities of the thyroid gland according to the EU-TIRADS system with data from image diagnostics based on artificial intelligence.
3. The proposed approach and classifier have been verified and validated based on program implementation and simulation experiments in .NET 4.6.1, .NET Core 2.2 and ML.NET 1.5.1 environment and using ResNet-50 deep neural network and the algorithm of Naive Bayes.

- V. **D. Ivanova**, B. Nenchovski, *Experimental Framework and Graphical User Interface for Research in Bioinformatics: GRAY – Gene Rapid Analysis*, DOI: 10.1109/InfoTech49733.2020.9210973, Conference: 2020 International Conference on Information Technologies (InfoTech), <https://ieeexplore.ieee.org/document/9210973>; [Scopus - Document details - Experimental Framework and Graphical User Interface for Research in Bioinformatics: GRAY - Gene Rapid Analysis](#)

The focus of this paper is on the numerical representation of gene data in bioinformatics. The benefits of using deep learning are presented, which are related to the fact that this type of algorithms complement DNA analysis, as they take into account individual variability in genes and can help in the treatment and prevention of diseases by determining the genetic predisposition to diseases.

The paper shows the process of creating a simple and convenient graphical user interface (GUI) that provides access to public biological databases, allowing the user to edit the data and analyze it with machine learning algorithms without requiring programming knowledge or specialized hardware.

Modern software libraries and the best programming practices are used to design the architecture and create the GUI.

Thus, the proposed experimental framework and graphical interface is aimed at researchers working with gene sequences and their analysis. The development presented in the paper offers an open-source software product "GRAY - Gene Rapid AnalySis".

This paper presents an experimental framework based on artificial neural networks such as the multilayer perceptron for the purpose of gene classification and analysis. It aims to reduce the time required to train such a network, in addition to lowering the hardware requirements for its virtualization without sacrificing accuracy.

### **CONTRIBUTIONS:**

1. A framework for analyzing genomic data with artificial intelligence is proposed.
2. An easy and convenient graphical user interface has been created that provides access to public biological databases and subsequent analysis with machine learning algorithms, without requiring programming knowledge or specialized hardware.
3. Open-source software "GRAY - Gene Rapid AnalySis" is developed for the big genomic data analytics.
4. The functionality of the proposed software product "GRAY" has been verified and validated based on experiments with different genomic data.

- VI. **D. Ivanova, J. Staeva, *Structural Similarity Analysis of SARS-CoV-2 and Human Proteins***, AIP Conference Proceedings 2505, 020001 (2022); <https://doi.org/10.1063/5.0100629>; [Scopus - Document details - Structural Similarity Analysis of SARS-CoV-2 and Human Proteins](#)

The purpose of this paper is to provide the structural similarity analysis of SARS-CoV-2 and human proteins. For this purpose, an experimental framework based on multiple innovative bioinformatics platforms is proposed.

This experiment aims to determine which human and SARS-2 proteins have the highest percentage of structural similarity and to measure this percentage. To accomplish this task, it is proposed a rather unconventional workflow in four phases: 1) Selection of a suitable database that contains human and SARS-2 proteins. 2) Phylogeny analysis to visualize the protein pairs with the lowest evolutionary distance. 3) Sequence alignment of the closest pairs of human and SARS-2 proteins using the phylogenetic tree results to measure their percent structural similarity. 4) Visualization and analysis of the results.

The results of the study are analysed in terms of the percentage of identity between human and SARS-2 protein pairs. The summary results show that the percentage of identity between human and SARS-2 protein pairs ranges from 7.6% to 23.6%. The degree of similarity varies between 8.5 % and 35.4 %. The protein pair with the highest identity and similarity score consists of proteins with shorter sequences.

### **CONTRIBUTIONS:**

1. A framework for structural similarity analysis of SARS-CoV-2 and human proteins is proposed.
2. A workflow covering all phases of the pipeline for structural similarity analysis of SARS-CoV-2 and human proteins is established.
3. The performance of the proposed framework for analyzing the structural similarity of SARS-CoV-2 and human proteins has been verified and validated based on numerous experiments conducted with different genomic data.

- VII. **D. Ivanova, V. Kadurin**, A new proposed software development methodology for healthcare industry, AIP Conference Proceedings 2333, 030010 (2021); <https://doi.org/10.1063/5.0042261>; [Scopus - Document details - A new proposed software development methodology for healthcare industry](#); [A New Proposed Software Development Methodology for Healthcare Industry-Web of Science Core Collection](#)

Digitizing healthcare systems and processes is essential to increase patient safety and reduce medical errors. The complex logic of healthcare and zero tolerance for errors requires special processes to create medical software. The purpose of this paper is to highlight and develop the main differences between all the significant and widely used software development methodologies, as they all have their useful properties and drawbacks in different situations. The paper discusses the general usability of the methodologies, including a description of how culture, continuous integration, and delivery are achieved with each one.

Successful implementation of healthcare software is a difficult task due to the nature of the domain. It has properties and functions that are very different from any other software product. Over the past two decades, healthcare applications and the use of information systems in the healthcare industry have grown rapidly. The main purpose of these applications is to improve the quality of medical care and service.

This paper proposes a new methodology for healthcare software development. It includes 4 distinct phases that are executed sequentially but have internal iterative patterns – 1) analysis plan, 2) architecture, 3) development, 4) validation and verification. This brings the benefits of both the waterfall and agile approaches. This combined approach specifically addresses the needs of the complex healthcare field.

### **CONTRIBUTIONS:**

1. An analysis of the scientific output related to modern achievements in the field of software engineering and software development methodologies was carried out by applying a bibliometric approach and a systematic literature review.
2. A new healthcare software development methodology is proposed, including 4 separate phases that are executed sequentially, but also have internal iterative models - analysis plan, architecture, development, validation, and verification.

**CONTRIBUTIONS from scientific publications that are referenced and indexed in world-renowned databases of scientific information (Web of Science and Scopus), outside of the habilitation work for a group of INDICATORS Γ**

The paper from the set of documents Γ1, Γ2, Γ3, Γ4, Γ5 are related to the modern achievements in the use of digital technologies for the interactive presentation of the Bulgarian cultural and historical heritage and more specifically: UI/UX design, programming for mobile devices and the creation of attractive mobile applications in the world of cultural and historical heritage.

**CONTRIBUTIONS:**

1. A conceptual model "Virtual piazza of the Bulgarian cultural and historical heritage" is proposed.
2. The scenario for virtual tours in the virtual piazza with a "Hop on, hop off" bus was created for an interactive presentation of the Bulgarian cultural and historical heritage using modern digital technologies with virtual and augmented reality.
3. A methodological process of mobile application development is proposed, reflecting the challenges and best practices in the field.
4. An architecture of a mobile application for interactive presentation of the Bulgarian cultural and historical heritage is proposed.
5. A mobile application for the Bulgarian cultural and historical heritage with gamification and geolocation has been developed, which offers users an interactive and convenient way to discover and get to know the rich history and culture of Bulgaria.
6. An analysis of the scientific production related to modern achievements in the field of UI/UX design was carried out by applying a bibliometric approach and a systematic literature review.
7. A usability strategy and guidelines for the development of a web portal for the Bulgarian cultural and historical heritage are proposed.
8. The hybrid approach is proposed, which combines proven and working methodologies and innovative processes focused on user-centered design.



9. An innovative user-oriented framework for the software development for the interactive presentation of the Bulgarian cultural and historical heritage is proposed.

The presented papers Г6 and Г7 are related to research in the field of the Internet of Things ecosystem, Big Data Analytics, and the application of Artificial Intelligence methods. Contributions can be formulated as follows:

### **CONTRIBUTIONS:**

1. A framework and architecture of an intelligent air quality assessment and monitoring system in IoT ecosystem is proposed.
2. A predictive model with machine learning based on multi-layer perceptron is created for air quality assessment.
3. A graphical user interface of the intelligent system has been developed to allow users to monitor the predicted air quality values.
4. A hybrid conceptual framework of the interaction between management control systems and big data analysis is proposed, which enables the successful achievement of the company's strategic goals and the assessment of the impact on the fourth levers of control.

### **CONTRIBUTIONS of a scientific publication with an impact factor (IF on Web of Science) and/or with an impact rank (SJR on Scopus), for a group of INDICATORS 3**

The papers in the document set 31, 32 and 33 are related to the security of the IoT ecosystem, the latest trends in IT and the latest advances in science in support of personalized medicine for breast cancer risk assessment and big genomic data analysis. The contributions are as follows:

### **CONTRIBUTIONS**

1. The conceptual model of a platform for intelligent security solutions in the IoT ecosystem is proposed.
2. A conceptual software architecture of an intelligent attack detection system is created, consisting of two sections: 1) for machine learning (ML) and 2) operational section.

3. An approach for building parallel differentiated diagnostic analytical workflows for the case of intrusion and attack detection is proposed.
4. Verification and validation of the proposed intelligent system is done based on a set of experiments for the diagnosis case.
5. The classification of breast cancer risk factors and the available risk assessment models and tools are proposed.
6. Breast cancer risk assessment models have been developed. A knowledge data discovery (KDD) bioinformatics workflow is established to assess breast cancer risk and determine risk type: high, elevated, and low risk.
7. Verification and validation of the developed models and the proposed bioinformatic workflow for breast cancer risk assessment based on experimental studies is done. The results obtained with an accuracy of up to 93% in the risk assessment of breast cancer prove the competitiveness of the proposed solutions.
8. An intelligent method for adaptive in silico knowledge discovery based on big genomic data analysis is proposed.
9. An architecture of the software system built based on the proposed method is designed.
10. Verification and validation of the proposed method is done for the case studies of differential descriptive analysis workflows for prokaryotic and eukaryotic gene discovery and mapping and differential diagnostic analysis workflows for breast cancer-associated gene mutation detection are performed.