

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ

«Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ» КЕАҚ

**Студенттер мен жас ғалымдардың
«GYLYM JÁNE BILIM - 2025»
XIX Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ**

**СБОРНИК МАТЕРИАЛОВ
XX Международной научной конференции
студентов и молодых ученых
«GYLYM JÁNE BILIM - 2025»**

**PROCEEDINGS
of the XX International Scientific Conference
for students and young scholars
«GYLYM JÁNE BILIM - 2025»**

**2025
Астана**

УДК 001(06)
ББК 72я631
F96

**«ǴYLYM JÁNE BILIM – 2025» студенттер мен жас ғалымдардың
XX Халықаралық ғылыми конференциясы = XX Международная
научная конференция студентов и молодых ученых «ǴYLYM JÁNE
BILIM – 2025» = The XX International Scientific Conference for
students and young scholars «ǴYLYM JÁNE BILIM – 2025». – Астана:
– 3813 б. - қазақша, орысша, ағылшынша.**

ISBN 978-601-08-5373-7

**Жинаққа студенттердің, магистранттардың, докторанттардың және жас
ғалымдардың жаратылыстану-техникалық және гуманитарлық ғылымдардың өзекті
мәселелері бойынша баяндамалары енгізілген.**

**The proceedings are the papers of students, undergraduates, doctoral students and young
researchers on topical issues of natural and technical sciences and humanities. В сборник
вошли доклады студентов, магистрантов, докторантов и молодых ученых по
актуальным вопросам естественно-технических и гуманитарных наук.**

УДК 001(06)
ББК 72я431
F96

ISBN 978-601-08-5373-7

Л.Н. Гумилев атындағы Еуразия
ұлттық университеті, 2025

		сауаттылығын арттыру	
203.	Эрболат А.	Орта мектепте нанотехнология ұғымын оқытудың тиімді әдістері	808

СЕКЦИЯ 2

СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

Подсекция 2.1			
Цифровая трансформация образования			
204.	Адалбек Н.	«Традиционные и интеллектуальные подходы в обучении»	812
205.	Бакенова А.А.	«Цифровизация тестирования: разработка нейросетевого приложения для формирования заданий по английской грамматике»	816
206.	Бекмурат А.Е.	«Инновационные методы обучения информатике в школе на основе искусственного интеллекта»	821
207.	Назарова А.Т.	«Развитие цифровых компетенций учителей в условиях персонализированного обучения»	826
208.	Нуриева Д.Р.	«Цифровая трансформация педагогики: роль информационных технологий в повышении квалификации преподавателей»	830
209.	Абдуашимова П.М.	«Білім беру процесінде жасанды интеллект технологияларын қолданудың тиімділігі»	833
210.	Ажибаева А.Д.	«Мектеп информатикасын оқытудағы кемшіліктерді жою жолдары»	837
211.	Асылбек М.А.	«Орта мектепте білім беру үдерісінде үлкен деректерді қолдану әдістемесі»	842
212.	Аталова А.Е.	«Әлеуметтік желілерді информатика пәні бойынша оқыту құралы ретінде пайдалану»	845
213.	Балтабаев Н.П.	«Мектептерде сабақ кестесін автоматтандыруға арналған интеллектуалды жүйе құру»	851
214.	Балтабаев Н.П., Дәрменов Ә.М., Мұратова М.М.	«Жасанды интеллект негізінде жаратылыстану пәндерін оқытуды жетілдіру: BilimALL AI платформасының мүмкіндіктері»	854
215.	Баумуратова Х.Б.	«АКТ оқыту барысында бастауыш сынып оқушыларының цифрлық сауаттылықтарын қалыптастырудың әдістемесі»	856
216.	Баумуратова Ш.Б.	«Жасанды интеллект негізінде инклюзивті білім беруді жетілдіру»	859
217.	Ғазиз Ж.Е.	«Бастауыш мектепте ақпараттық-коммуникациялық технологияларды оқыту әдістемесі»	863
218.	Дәрменов Ә.М.	«Информатиканы қолжетімді ететін мобильді "BilimAll" қосымшасы»	866
219.	Дүйсегалиева Н.А.	«HIGH-TOUCH HIGH-TECH моделі арқылы болашақ информатика мұғалімдерін машиналық оқыту негізінде даярлаудың	870

	инновациялық тәсілдері туралы»	
220.	Еликбай А.Ж. «Ақпараттық дәуірде білім берудің жаңа кезеңі – Инфографика»	874
221.	Жаңабекқызы А. «EDCAFE AI көмегімен сабақты жоспарлау»	879
222.	Жумабекова У.Б., Сабырова М.Е., Сабыров Т.С. «Информатика пәнін жобалап оқыту технологиясы»	883
223.	Кендебай Н.А. «EDUVISION білім беру процесін қадағалайтын қосымша»	888
224.	Көшенова А. «Цифрлық сауаттылықтың мектеп курсы бойынша интеллектуалдық оқу басылымдарына арналған дидактикалық материалдар»	891
225.	Куанышева Д.Ж. «Инклюзивті білім беруде педагогтың ақпараттық-коммуникациялық технологияларды (АКТ) қолдану даярлығын жетілдіру»	893
226.	Мауленова М.А. «Үлкен деректерді өңдеуде машиналық оқытудың әдістері мен құралдары»	897
227.	Мылтыкбаева Ж.Т. «Жаратылыстану пәндерін STEM білім беру мен ROS операциялық жүйесі негізінде кешенді оқыту»	901
228.	Надирхан Г.Е. «Ауыл мектептерінде цифрлық оқытуды дамыту мүмкіндіктері»	903
229.	Орынбаев М.Ж. «Компьютерлік көру алгоритмдерін машиналық оқыту негіздері бойынша қолданудың оқу-әдістемелік негіздері»	907
230.	Сабитова А.Б., Ражапова А.Н. «Жасанды интеллект және білім: болашақ мұғалімдерге арналған жаңа мүмкіндіктер»	910
231.	Сағындықова А.С. «Болашақ информатика мұғалімдерін магистратураға даярлаудағы онлайн-курстардың рөлі»	915
232.	Сайлау Ж.Б. «Халықаралық зерттеуге оқушыларды АКТ арқылы дайындаудағы педагогтердің құзыреттілігін арттыру жолдары»	918
233.	Төрәлі Қ.Н. «Бастауыш сынып оқушыларының цифрлық сауаттылығын дамытудың ерекшеліктері»	923
234.	Турмаганбетова З.П., Алтыбаева А.Н. «Ерекше білімді қажет ететін оқушыларға мектеп информатика курсы оқытуды ұйымдастыру»	927
235.	Халхабай А. ««Алгоритмдеу және бағдарламалау» курсы бойынша мобильді қосымшаны оқу үдерісінде қолдану»	931
236.	Ысмайыл Н. «Мектеп информатика курсына жобалық оқыту әдісін енгізу»	936
237.	Ізбасарова М.Р. «Білім берудегі тестілеу жүйелері»	938

Подсекция 2.2

Интеллектуальные информационные системы

238.	Amantayeva Gulden Turarkyzy «Comparative analysis of models and methods in heart disease prediction problems»	944
------	---	-----

239.	Tanirbergenov Meirbek Sagyndykovich «Facial Recognition-Based Attendance Management»	947
240.	Toleubay Daniyar Manatuly «Cardiac disease prediction using machine learning algoritms»	952
241.	Yerezhepov Rakhat Aibulatovich «Detecting logical fallacies in web content with nlp-powered crawling»	957
242.	Ажикенов Арман Русланович, Абашев Арслан Азатабекович «Оптимизация дорожного трафика в Астане через симуляцию транспортных потоков»	962
243.	Аманжол Альфараби Маликович, Сабит Мадияр, Кушербаев Бекзат Алибекулы «Система визуализации и анализа данных о передвижении нефти на основе интерактивной карты»	968
244.	Аскапова Мадина Куанышбековна «Параллельді қазақ-түрік сөйлеу корпусы қалыптастырудың әдісі мен моделін құру»	972
245.	Бекқожин Дастан Ақанұлы «Терең оқыту негізінде қолжазба таңбаларын тану программалық құралын әзірлеу»	975
246.	Дакенов Алишер Мырзахметұлы «Анализ сигналов ЭЭГ нейросетевыми методами для ранней диагностики нейродегенеративных заболеваний»	978
247.	Доспол Нәзгүл Нурланқызы, Жеткенбай Лена «Балабақшадағы балалардың эмоциялық жағдайын бақылауға арналған эмоцияларды тану жүйесін әзірлеу»	987
248.	Ермекбай Айболат, Молдабек Елжан «Жасанды интеллект негізінде веб-қосымша әзірлеу»	992
249.	Жұмал Жания Ержанқызы, Абдурахман Жансая Берікжанқызы «Применение голосового ИИ-помощника в геймифицированной образовательной среде»	1001
250.	Каримов Руслан Жасинович «Эффективность существующих ИИ-решений в основных направлениях транспортной логистики»	1007
251.	Кубиева Сабина Талгатовна, Утепбергенова Зарина Арманкызы «Разработка iot системы по уходу за растениями на базе искусственного интеллекта»	1012
252.	Кудобаев Даниал Дулатович «Разработка информационной системы для автоматизации стоматологических услуг»	1017
253.	Мусина Данель Тлеухановна «Интеллектуальные инструменты автоматизированной диагностики надежности информационных систем»	1024
254.	Рогова Ксения Александровна, Қабдыбек Ризат Досмжанұлы, Джумадиева Тогжан Бекежановна «Мониторинг инженерных конструкций на основе искусственного интеллекта»	1030

255.	Сафонова Софья Александровна «Современные аспекты информационной безопасности в облачных вычислениях: модели, угрозы и методы защиты»	1034
256.	Смаилова Назгүл Батырбекқызы «Терең оқыту арқылы кітап ұсыныстарын әзірлеу: collaborative filtering, content-based және nlp әдістерінің комбинациясы»	1041
257.	Тажібай Аружан Айдосқызы, Кудубаева Сауле Альжановна «Көру қабілеті әлсіз адамдарға арналған ai дауыстық көмекші: нақты уақытта объектілерді анықтау және қашықтықты бағалау»	1046
258.	Тайжанов Азамат Жанкелдіұлы «Python тілінде фильмдердің интеллектуалды ұсыныс жүйесін әзірлеу»	1051
259.	Умирзахов Сундетали Кабылбекович «Сұраныстарды интеллектуалды талдау негізінде ұйымның сайты үшін чат-бот құру»	1055
260.	Шайхстан Марғұлан «ИОТ Сенсорлары негізінде ауа ластану деңгейін болжау»	1060

Подсекция 2.3

Современные тенденции в программной инженерии и управлении в условиях цифровой индустрии

261.	Bekenova A.B. «Development of a registration panel for users and doctors with integration into the database»	1077
262.	Bolat A.Zh. «Data analysis methods and decision making using big data and machine learning tools»	1081
263.	Алтайұлы А. «Visual studio интегралды ортасында «қойма қызметкерлеріне арналған» мәліметтер қорын жобалау»	1086
264.	Арап А.Қ. «Ақылды сурет салушы роботты әзірлеу»	1088
265.	Артыкбекқызы А. «Ақылды үйлердегі заттар интернеті(iot) мен робототехниканың өзара әрекеттесуі»	1091
266.	Ахметова А.Д. «Тоңазытқыштағы өнімдерді бақылауға және тағам әзірлеу ұсынысын беруге арналған программалық қосымша»	1096
267.	Дәрібай Д.Д. «Робототехниканы қолдану арқылы қойма логистикасындағы қолданыстағы басқару жүйелерін талдау»	1100
268.	Жамбулов С.Ж. «Білім алушыларды информатика және программалау олимпиадаларына дайындауда жасанды интеллекттің қолданысы»	1102
269.	Каиржан Р.С. «Development of system for recognition of emotional states of employees based on computer vision methods on Raspberry Pi»	1108
270.	Кайрекенова Н.Р. «Өнеркәсіптік роботты көру үшін машиналық оқытудың заманауи тәсілдері: әдістер, деректер жиынтығы және оптимизациялау»	1111

271.	Калижан А.К. «Разработка системы биометрической аутентификации с предотвращением deepfake атак»	1113
272.	Касылкасова К.Н. «Программное обеспечение smartmed для обработки медицинских данных и диагностики»	1118
273.	Қабдешев Ә.Е. «Жөтелді талдау негізінде денсаулықты диагностикалаудың интеллектуалды программасын әзірлеу»	1120
274.	Махаев Е.Е. «Разработка облачного приложения для автоматизации деятельности сети аптек»	1123
275.	Муратов М.М. «Эффективность единой информационной системы агропромышленного комплекса»	1126
276.	Нуржанова А.Б. «Современные методы классификации эмоций: анализ подходов и перспективы развития»	1130
277.	Нурпеисова З.Р. «Обзор и исследование методов искусственного интеллекта для анализа рынка недвижимости»	1134
278.	Рақымбек А.С. «Кітапқұмарларға арналған платформа: кітаптарды оқу және бөлісу үшін әлеуметтік желіні жобалау және іске асыру»	1138
279.	Сагидуллина Д.С. «Visual studio интегралды ортасында «қаржылық транзакцияларды қадағалау және талдауға арналған» мәліметтер қорын жобалау»	1144
280.	Төлеубай Д.М. «Yolov10 қолдану арқылы рентген суреттерінде сүйек сынуын анықтауды кешенді зерттеу»	1147
281.	Утегенова Д.Б. «Visual studio интегралды ортасында «фитнес орталық қызметкері үшін» мәліметтер қорын жобалау»	1152
282.	Шаймуратов А.Ж. «Проектирование аппаратно-программного комплекса для автоматизированного учета железнодорожного подвижного состава»	1154
Подсекция 2.4		
Информационная безопасность		
283.	Akniyet N. «Smart home automation and security system using arduino uno r4 and esp32 microcontrollers with telegram integration»	1158
284.	Askhatov A. «Analysis of social engineering methods and development of a defense strategy for corporate structures»	1165
285.	Bekturganov A.B. «Development of an early detection model for ddos attacks based on network traffic analysis»	1170
286.	Gabdullin A. «Analysis of modern wireless network security protocols and prospects for their development»	1174

287.	Garifullin A. «Modern information security management systems: construction and implementation in the digital era»	1179
288.	Igumenshev D.V. «Methods of embedding malicious code into pdf files»	1182
289.	Issabay T.B. «Utilizing sandboxes for cybersecurity training: a hands-on approach»	1187
290.	Kalybayev S. «Overview of modern authentication methods in telecommunication systems: from passwords to biometrics»	1191
291.	Kerim A. «Owasp top 10 and alternative methods of its compilation»	1194
292.	Yergazin A. «Analysis of a protection of hybrid intrusion detection and prevention system (idps) for low-latency 5g networks with adaptive learning using edge computing»	1199
293.	Yerzhanova Y.Y. «Key attacks in web forensics: xss, sql injection and rce»	1204
294.	Zhakay A. «Fundamentals of modern cryptography: from encryption to digital signatures»	1209
295.	Айдарова А.А. «Visualvm көмегімен cast-128 және kuznyechik блоктық шифрларының кілт генерациясын салыстыру және стандарттарға шолу»	1214
296.	Акимбекова Д.М., Каиржанова Д.Ж. «Жергілікті желінің қауіпсіздігін қамтамасыз ететін негізгі параметрлер»	1220
297.	Аскарлов А.Д. «Разработка и исследование эффективности метода и инструмента для выявления фейковых новостей в социальных сетях»	1224
298.	Ауесхан Н. «Аномалияларды анықтау әдістерін талдау»	1229
299.	Ерболатов А. «Анализ вредоносных программ с помощью ии и криптографическая защита»	1332
300.	Ерболатова А.Ж. «Neuvector және kubernetes: контейнерлік ортадағы қауіпсіздікті қамтамасыз ету тәсілдері»	1336
301.	Жанатаев М.К. «Стеганография на основе lsb: реализация сокрытия данных в медиафайлах»	1338
302.	Жарасхан Н.Ж., Қайупов Е.К. «Crystals-kyber алгоритмін ресурсы шектеулі құрылғыларға оңтайландыру»	1343
303.	Жолдасбаев М.Ә. «Заманауи операциялық жүйелердегі жады дампы кескінін алу құралдарын талдау және салыстыру»	1348
304.	Жолмұратұлы Б., Маратов Ә.Б., Ховдабай Н.А. «Екі факторлы	1353

	аутентификацияның қауіпсіздігі және оның қолданылуы»	
305.	Кадринов Д.М. «Автоматизация внедрения альтернативной soag платформы на основе средств со свободной лицензией»	1357
306.	Казбаганбетова М.А. «Wireshark бағдарламасын пайдаланып желілік трафикті талдау және ақпараттық қауіпсіздікті қамтамасыз ету»	1361
307.	Кәкімбек Ә.Қ., Серікбай А.Е., Наурызбаев Д.Е. «MITM шабуылы туралы»	1366
308.	Кеттеш Б.Н. «ELF талдауындағы capstone: сызықтық және рекурсивті дизассемблерлеу»	1370
309.	Көшкінбаева Ф.Қ. «Linux қорғаудың заманауи әдістеріне талдау.openvas және nmap көмегімен осалдықтарды анықтау»	1374
310.	Қадыр Н.Е. «Заманауи фишинг түрлері мен олардың ұйымдық ақпараттық жүйелерге ықпалы»	1379
311.	Қажкен Е.Е., Темиржан С.А. «Қауіпсіздік инциденттеріне қалай жауап беруге болады?»	1384
312.	Қартбай Е.Ғ., Тынарбай Н.И. «MITM шабуылы (адамның ортадағы шабуылы)»	1388
313.	Маратов Б.Ж. «Әлеуметтік инженерия қауіпсіздікке қатер ретінде: қызметкерлерді қорғау және оқыту әдістері»	1393
314.	Мағзумов А.М. «Websocket протоколындағы осалдықтарды талдау»	1397
315.	Майданов А.С. «Автоматизация процесса анализа оперативной памяти с использованием python»	1401
316.	Мақсат Ә., Нурсейтов С. «Блокчейн қажеттілік пе, әлде сән бе?»	1406
317.	Қ. Мырзағалиұлы. «Инциденттерді анықтауда желілік логтарды талдаудың маңызды рөлдері»	1409
318.	Нурбатуров С.К. «Интеграция honeypot в ит-инфраструктуру компании»	1412
319.	Нуриева Д.Р., Исайнова А.Н. «Анализ рисками безопасности данных в медицинских учреждениях»	1415
320.	Нұрлан А.Т. «Кескіндердегі статистикалық стегоанализ әдістері»	1420
321.	Оралбеков Е.А. «Ddos-шабуылдардың жаңа буыны»	1424

322.	Рамазанова Ж, Нұрлан А, Жайсанбаева А. «Бұлтты технологияларды пайдалану кезіндегі тәуекелдер мен қауіпсіздік шараларын зерттеу»	1430
323.	Сахатбекқызы Т., Бахтиярқызы Т.А. «IoT құрылғыларының желідегі қауіпсіздігін қалай қамтамасыз етуге болады: стратегиялар және packet tracer көмегімен модельдеу»	1434
324.	Серғазы М. «Повышение производительности разработчиков с помощью интегрированных искусственных интеллектов и соображения кибербезопасности»	1440
325.	Султанов А.М. «Стеганография в кибербезопасности казахстана»	1443
326.	Танатаров Е., Іргебай С., Султанов А. «WI-FI желісінде шақырылмаған қонақтарды автоматты түрде анықтау жүйесі»	1447
327.	Таубай М.Е. Раматуллаев Ә.А. «Фишинг: желідегі beef әдісі арқылы алдау және одан сақтану»	1452

СЕКЦИЯ 3 ЕСТЕСТВЕННЫЕ НАУКИ

		ПОДСЕКЦИЯ 3.1 АКТУАЛЬНЫЕ ПРОБЛЕМЫ СОВРЕМЕННОЙ БИОЛОГИИ	
328.	Акимкара А.Б.	Гербарийдің ботаникалық зерттеулерде қолданылуы және гербарий қорындағы кеппе шөптің қалыптасу ерекшеліктері	1457
329.	Ақылбек А.	Астана қаласындағы ботаникалық бағының ландшафттағы <i>geranium sanguineum</i> биологиялық ерекшеліктеріне сипаттама беру	1459
330.	Әділхан Ж.	Мобильді байланыс пен қолданбалардың адамның мінез-құлқына әсерін анықтау	1463
331.	Базарбаева Қ.	Жасөспірімдерде девиантты мінез-құлықтың даму қаупі	1467
332.	Байдосова А.Б.	Методика использования игровых технологий на уроках биологии	1471
333.	Байдосова А.Б.	Актуальные проблемы современной биологии с использованием игровых технологий в образовании	1474
334.	Ғазизова Ә.	Сәулеленген егеуқұйрықтардың бүйректеріндегі морфофункционалдық өзгерістерді салыстырмалы бағалау	1477
335.	Еркін З.Б.	Биология сабақтарында оқушылардың сыни ойлау қабілетін жетілдіруде блум таксономиясын пайдалану	1482
336.	Жанабергенова	Кенеттен жүрек өлімі: генетикалық аспектілері	1486

CARDIAC DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS

Toleubay Daniyar Manatuly

daniyar_manatuly@mail.ru

Student of L.N. Gumilyov Eurasian National University

Department of «Information systems»

Supervisor – Prof. Dr. Zhukabayeva Tamara Kokenovna

Abstract

This study focuses on the application of machine learning methods for predicting cardiac diseases, aiming to develop accurate and interpretable diagnostic models. The primary objective is to enhance prediction accuracy and identify key risk factors. Various classification algorithms, including Random Forest, Gradient Boosting, Support Vector Machines (SVM), and Neural Networks, are utilized. A dataset containing clinical attributes of patients is processed and analyzed to build predictive models. The model performance is evaluated using accuracy metrics, confusion matrices, and ROC-AUC scores. The results indicate that the Random Forest algorithm achieves the highest predictive accuracy. Furthermore, feature importance analysis highlights the most significant factors contributing to the development of cardiac diseases. This research provides valuable insights for medical professionals and researchers, offering effective tools for early disease detection.

Key words: Cardiac disease, Machine Learning, Diagnosis, Classification, Data Analysis, Random Forest, Gradient Boosting, SVM, Neural Networks.

Introduction

Cardiac disease is one of the leading causes of mortality worldwide, placing a significant burden on healthcare systems. Early diagnosis plays a crucial role in reducing mortality rates and improving treatment effectiveness. However, traditional diagnostic methods often rely on costly laboratory tests and instrumental examinations, making timely identification of high-risk patients challenging.

In recent years, machine learning has been widely applied in the medical field, providing new opportunities for analyzing large volumes of clinical data and improving prediction accuracy. Various classification algorithms, such as Random Forest, Gradient Boosting, and Neural Networks, have demonstrated their effectiveness in disease diagnosis.

The goal of this study is to develop and analyze machine learning models for predicting cardiac disease. The research involves data preprocessing, feature selection, model training, and evaluation using multiple classification algorithms. The models are assessed based on prediction accuracy, confusion matrices, and ROC-AUC scores. Additionally, feature importance analysis is conducted to identify the most critical clinical indicators influencing disease development.

Dataset description

The dataset used in this study originates from the UCI Heart Disease dataset, which contains clinical and diagnostic attributes of patients. It includes 303 observations with 14 features related to medical conditions and risk factors. The target variable indicates the presence (1) or absence (0) of cardiac disease.

The key attributes include age, sex, chest pain type (cp), resting blood pressure (trestbps), serum cholesterol (chol), fasting blood sugar (fbs), resting electrocardiographic results (restecg), maximum heart rate achieved (thalach), exercise-induced angina (exang), ST depression induced by exercise (oldpeak), slope of ST segment (slope), number of major vessels colored by fluoroscopy (ca), and thalassemia type (thal).

This dataset is widely used in machine learning research for cardiac disease prediction, as it provides a mix of categorical and continuous variables that influence heart health. The diversity of features enables the development of predictive models that assist in early disease detection and risk assessment.

Pretrained model

Before applying machine learning algorithms, data preprocessing was conducted to improve model performance. The target variable was transformed into a binary format, where 0 indicates the absence of cardiac disease and 1 represents its presence. The dataset was checked for missing values, but no significant gaps were found. Features were standardized using StandardScaler to ensure uniform scaling, which enhances model convergence. Categorical variables were encoded to numerical values where necessary. The dataset was then split into training (80%) and testing (20%) sets using `train_test_split()` to ensure reliable model evaluation. These preprocessing steps help optimize the learning process, improve classification accuracy, and enhance the interpretability of the results.

Training process

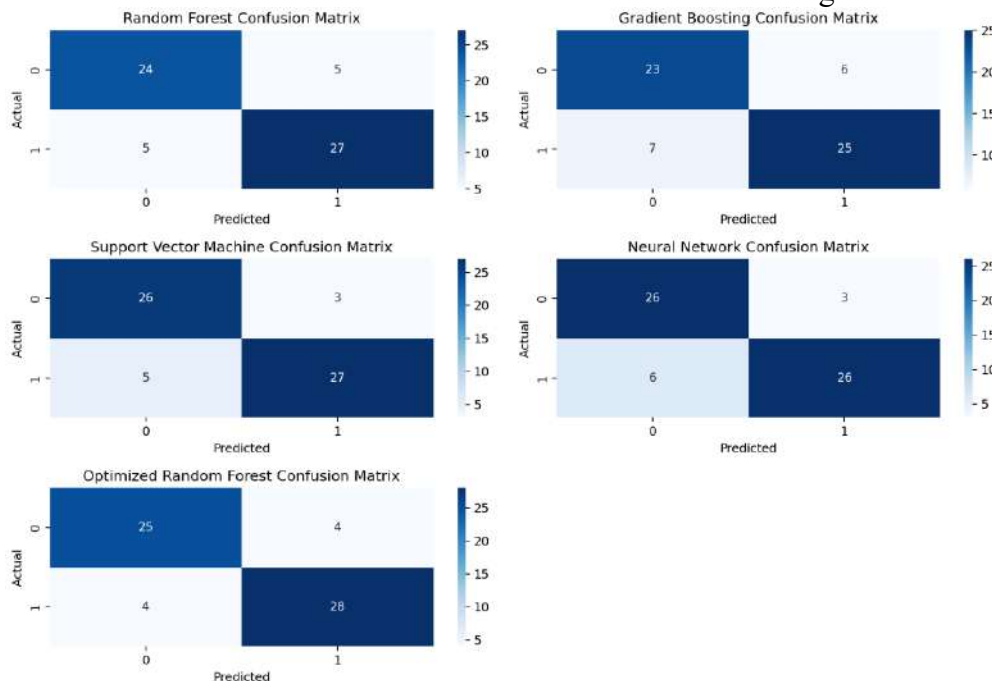
Four machine learning models – Random Forest, Gradient Boosting, Support Vector Machine (SVM), and Neural Network (MLPClassifier) – were selected for training. To enhance performance, hyperparameter tuning was conducted using GridSearchCV for the Random Forest model, optimizing the number of estimators and tree depth. The batch training strategy and cross-validation were employed to improve model robustness and prevent overfitting.

Each model was trained and evaluated based on accuracy, confusion matrices, classification reports, and ROC-AUC scores. The Random Forest model demonstrated superior performance, achieving the highest predictive accuracy. A feature importance analysis was conducted to identify key clinical attributes influencing cardiac disease development.

Throughout training, model performance was continuously monitored, adjusting hyperparameters when necessary. The evaluation results confirmed that machine learning techniques provide an effective approach for early cardiac disease detection, with Random Forest proving to be the most reliable classifier.

Interpretation of Results

Figure 1. Confusion matrix.



The performance of each model was assessed using confusion matrices, which provide insight into prediction accuracy across different classes. The confusion matrix evaluates the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), helping to understand misclassification patterns.

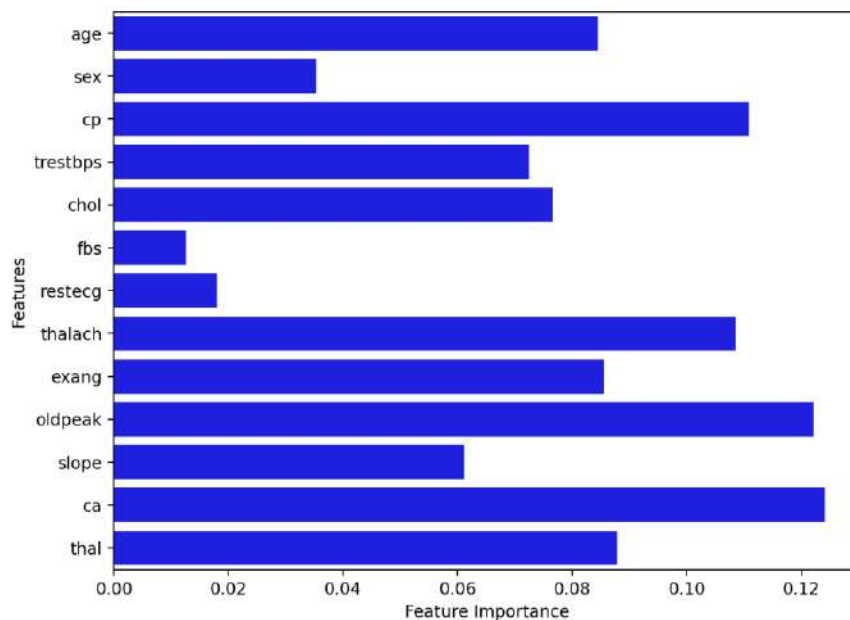
The Random Forest model demonstrated strong classification capabilities, with a high number of correct classifications in both positive and negative cases. The Optimized Random Forest model further improved the results by reducing false positives and false negatives through hyperparameter

tuning. Gradient Boosting and Support Vector Machine (SVM) also achieved competitive performance, though they exhibited slightly higher misclassification rates.

The Neural Network (MLPClassifier) showed some variability, with more false negatives compared to tree-based models, indicating potential challenges in recognizing certain cases. The heatmaps of confusion matrices visually highlighted these differences, confirming that Random Forest-based approaches provided the most balanced and reliable classification for heart disease prediction.

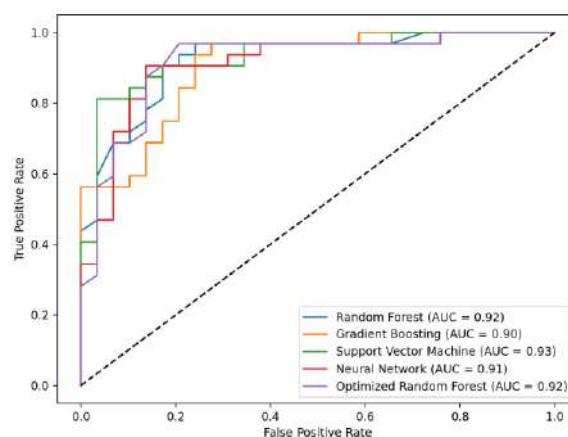
To enhance the interpretability of the model, feature importance analysis was conducted using the best-performing classifier, the Optimized Random Forest. Feature importance provides insights into which clinical attributes have the most significant impact on predicting cardiac disease. The Random Forest algorithm assigns an important score to each feature based on how much it contributes to reducing impurity in the decision trees. The analysis revealed that attributes such as thalassemia type (thal), number of major vessels (ca), maximum heart rate achieved (thalach), and ST depression induced by exercise (oldpeak) had the highest influence on the model's predictions. A visualization of the feature importance scores is presented in Figure 2, which demonstrates the relative contribution of each clinical attribute.

Figure 2. Feature importance in optimized Random Forest



The Receiver Operating Characteristic (ROC) curve was used to evaluate the models' ability to distinguish between positive and negative cases. The AUC (Area Under the Curve) metric provides a numerical measure of this capability, with higher values indicating better performance.

Figure 3. ROC curves.



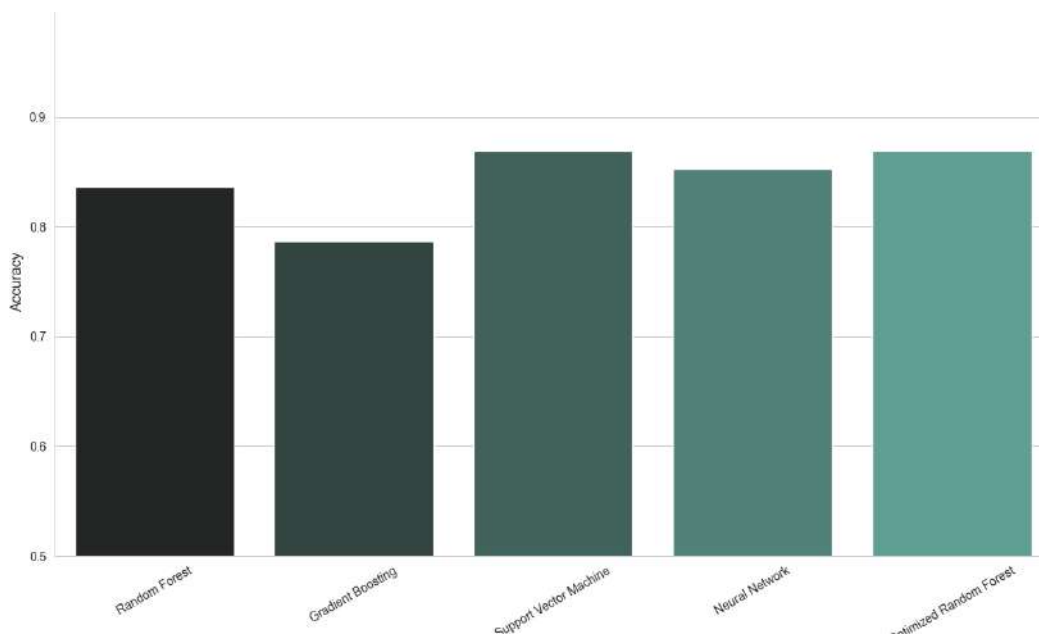
Among the evaluated models, the Support Vector Machine (SVM) demonstrated the highest AUC score of 93%, indicating its superior ability to differentiate between patients with and without heart disease. The Optimized Random Forest, Gradient Boosting, and Neural Network models also achieved competitive AUC scores, further validating their effectiveness. As shown in Figure 3, the ROC curves illustrate the models' classification performance across different thresholds. A steeper curve closer to the top-left corner represents a more effective model, reinforcing the SVM model's robustness in this context.

Figure 4 presents a comparison of the accuracy of various machine learning models. The best-performing models were Support Vector Machine (SVM) and Optimized Random Forest, both achieving an accuracy of 87%.

Random Forest and Neural Network showed similar results with accuracies of 84% and 85%, respectively. Meanwhile, Gradient Boosting had the lowest accuracy at 79%, which may be due to the characteristics of the dataset and the chosen hyperparameters.

These results indicate the high efficiency of the SVM and optimized Random Forest methods for this classification task.

Figure 4. Model performance comparison.



The selection of machine learning models was based on their proven effectiveness in medical diagnosis tasks. Support Vector Machine (SVM) is widely used in healthcare due to its ability to handle high-dimensional data and find optimal decision boundaries, making it particularly suitable for disease classification problems [6]. Random Forest is a robust ensemble method that reduces overfitting and can efficiently process both numerical and categorical features, which is beneficial for analyzing clinical datasets [7]. Optimized Random Forest, through hyperparameter tuning, further improves the classification performance by minimizing false positives and false negatives. Gradient Boosting, particularly in implementations like XGBoost, is known for its high predictive accuracy and has been successfully applied in disease risk assessment tasks [8]. However, its sensitivity to hyperparameter selection can sometimes lead to overfitting or suboptimal results. Neural Networks excel at capturing complex nonlinear patterns in medical data, making them effective in disease prediction models [7]. Although they require larger datasets and extensive tuning, they are valuable for identifying intricate relationships between clinical features.

Conclusion

The comparison of machine learning models for heart disease classification highlights the effectiveness of Support Vector Machine (SVM) and Optimized Random Forest, which demonstrated the highest accuracy (87%) and the best overall performance. These models exhibited strong

predictive capabilities, making them optimal choices for this classification task. The performance of Random Forest and Neural Network also showed promising results, indicating their potential with further tuning. However, Gradient Boosting showed the lowest accuracy, suggesting that additional optimization or feature selection may be needed to enhance its performance.

Machine learning models play a crucial role in predictive healthcare analytics, enabling the identification of patients at risk for heart disease. The findings suggest that SVM and Optimized Random Forest can be used as reliable tools for medical decision-making, potentially aiding clinicians in early diagnosis and risk assessment. Further research could explore additional feature selection methods and model ensembling techniques to further enhance classification performance.

Overall, these findings emphasize the importance of model selection, hyperparameter tuning, and feature importance analysis in achieving optimal results for heart disease prediction.

List of references

1. UCI Machine Learning Repository: Heart Disease Data Set. Website: <https://archive.ics.uci.edu/ml/datasets/heart+disease>.
2. Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists. O'Reilly Media, Inc.
3. Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). Foundations of machine learning. MIT Press.
4. Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 7, 81542-81554.
5. Kanamarlapudi, S., Yakkala, V. S., Gayathri, B., Nusimala, K. V., Aravinth, S. S., & S, S. (2023, March). Comparison and analysis of various machine learning algorithms for disease prediction. In *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 246-250). IEEE.
6. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). An overview on the advancements of support vector machines for medical data analysis. *Information*, 12(6), 219.
7. Wu, Y., Zhang, Y., Li, Z., & Wang, H. (2023). Heart disease detection using machine learning methods. *Journal of Medical Artificial Intelligence*, 6, 19.
8. Bhuiyan, M. A., Hossain, M. S., Muhammad, G., & Alamri, A. (2023). A systematic review of machine learning approaches in diabetes management and complications. *Healthcare*, 11(2), 220.