

**МЕДИЦИНСКИ УНИВЕРСИТЕТ – СОФИЯ**  
**МЕДИЦИНСКИ ФАКУЛТЕТ**  
**КАТЕДРА ПО БЕЛОДРОБНИ БОЛЕСТИ**

---

Доц. д-р Евгени Витков Меков, дм

**ИЗКУСТВЕН ИНТЕЛЕКТ И МАШИНО  
ОБУЧЕНИЕ ПРИ ПРОГНОЗИРАНЕ НА  
ЧЕСТОТА НА ЕКЗАЦЕРБАЦИИ И  
ЕДНОГОДИШНА СМЪРТНОСТ ПРИ  
ПАЦИЕНТИ С ХОББ**

ДИСЕРТАЦИОНЕН ТРУД

за придобиване на научна степен „Доктор на науките“

Научна специалност – Белодробни болести

София, 2024 г.

Дисертационният труд съдържа 328 страници и е онагледен с 139 фигури и 80 таблици. Библиографията обхваща 484 литературни източника.

Дисертационният труд е обсъден и насочен за публична защита от Катедрен съвет на Катедра по белодробни болести, МФ, МУ-София на 13.06.2024 г.

Официалната защита на дисертационния труд ще се състои на 22.10.2024 г. от 13:30 ч. в аудитория №3 на УМБАЛ „Свети Иван Рилски“ ЕАД, гр. София, бул. „Акад. Иван Гешов” №19 пред научно жури в състав:

1. Проф. д-р Росен Емилов Петков, дм
2. Проф. д-р Димитър Темелков Костадинов, дм
3. Проф. д-р Венцислава Пенчева Пенчева-Генова, дм
4. Проф. д-р Благой Иванов Маринов, дм
5. Проф. д-р Димо Митев Димов, дм
6. Доц. д-р Динко Генчев Вълев, дм
7. Доц. д-р Милена Димитрова Енчева, дм

Материалите по защитата на дисертационния труд са на разположение в КББ, УМБАЛ „Св. Иван Рилски” ЕАД, бул. „Акад. Иван Гешов” №19, както и на интернет страницата на МУ-София: <http://career.mu-sofia.bg/>.

## Съдържание

Списък на използваните съкращения .....	4
1. <b>Въведение</b> .....	7
2. <b>Цел и задачи</b> .....	9
3. <b>Материали и методи</b> .....	10
4. <b>Резултати</b> .....	22
4.1. Дескриптивна статистика .....	22
4.2. Машинно обучение за прогнозиране на екзацербации ....	52
4.3. Риск по GOLD .....	62
4.4. Честота на екзацербации (непрекъсната променлива) .....	73
4.5. Честота на екзацербации (категорийна променлива) .....	80
4.6. Фенотипизиране .....	90
5. <b>Дискусия</b> .....	104
5.1. Избор на най-добър модел за прогнозиране наличие на екзацербация .....	114
5.2. Избор на най-добър модел за прогнозиране на риск по GOLD .....	114
5.3. Избор на най-добър модел за честота на екзацербации (непрекъсната променлива) .....	115
5.4. Избор на най-добър модел за честота на екзацербации (категорийна променлива) .....	115
6. <b>Заклучение</b> .....	117
Приноси .....	118
ПУБЛИКАЦИИ И УЧАСТИЯ В НАУЧНИ ФОРУМИ ПО ТЕМАТА НА ДИСЕРТАЦИЯТА .....	119

## **Списък на използваните съкращения**

### На български език:

БЕК - брой на еозинофили в кръвта  
БТЕ – белодробна тромбемболия  
ДК – дясна камера  
ДО - дихателен обем  
ЕКГ – електрокардиография, електрокардиографски  
ЗСЛК - задна стена на лява камера  
ИИ – изкуствен интелект  
ИКС – инхалаторни кортикостероиди  
ИТМ - индекс на телесна маса  
КС – кортикостероиди  
КТ – компютърна томография  
ЛК - лява камера  
МНМК - метод на най-малките квадрати  
МО – машинно обучение  
ОКС – орални кортикостероиди  
ОО – остатъчен обем  
ПКК - пълна кръвна картина  
ПХ – пулмонална хипертония  
СН – сърдечна недостатъчност  
ТДР - теледиастолен размер  
ТСР - телесистолен размер  
ФВК – форсиран витален капацитет  
ФЕО1 – форсиран експираторен обем за първата секунда  
ФИ - фракция на изтласкване  
ФИД - функционално изследване на дишането  
ФОК - функционален остатъчен капацитет  
ФС - фракция на съкращение  
ХОББ – хронична обструктивна белодробна болест

### На английски език:

6MWT - шестминутен тест с ходене  
А – А вълна (ехокардиография)  
AATD -  $\alpha$ 1-антитрипсинов дефицит  
АСО – астма-ХОББ припокриване  
ВМІ - индекс на телесна маса  
САТ - тест за оценка на ХОББ  
СРЕТ - кардиопулмонален тест с натоварване

CRP - С-реактивен протеин  
CRQ - въпросник при хронични белодробни заболявания  
DT – децелерационно време  
E – E вълна (ехокардиография)  
EDV - краен диастолен обем  
EF - фракция на изтласкване  
ELR - съотношение еозинофили/лимфоцити  
ESV – краен систолен обем  
ESWT - совалков тест за издръжливост с ходене  
ET-1 - ендотелин 1  
GOLD - Глобална инициатива за хронична обструктивна белодробна болест  
HPV - хипоксична пулмонална вазоконстрикция  
HRQoL - качество на живот, свързано със здравето  
EBR - съотношение еозинофили/базофили  
EHR - електронни здравни досиета (Electronic Health Records)  
ICS - инхалаторни кортикостероиди  
IL – интерлевкин  
ISWT - инкрементален совалков тест с ходене  
KNN - метод на k-най-близките съседи (K-nearest neighbors)  
LABA – дългодействащи бета2-агонисти  
LAMA – дългодействащи мускаринови антагонисти  
LAP - налягане в ляво предсърдие  
LVSD - систолна дисфункция на лявата камера  
MCS – оценка на менталния компонент  
MPV - среден тромбоцитен обем  
NLR – отношение неутрофили/лимфоцити  
NO - азотен оксид  
PAPm - средно налягане в пулмоналната артерия  
PCA - метод на главните елементи (principal components analysis)  
PCS – оценка на физическия компонент  
PDW – ширина на тромбоцитното разпределение  
PEF - върхов експираторен дебит  
PGI2 - простагландин I2, простаглицлин  
PLR - съотношение тромбоцити/лимфоцити  
PVR - белодробна съдова резистентност  
QTcd - дисперсия на коригирания QT интервал  
QTd - QT продължителност  
RDW – ширина на еритроцитното разпределение  
RMSE - корен от средната квадратична грешка

ROC - площ под кривата на работната характеристика  
RPA - диаметър на дясна пулмонална артерия  
S' - систолна скорост на латералния трикуспидален пръстен  
SF-36 - кратък въпросник с 36 елемента  
SGRQ - въпросник на болница Сейнт Джордж  
SpO<sub>2</sub> - кислородна сатурация  
SVM - поддържаща векторна машина (Support Vector Machine)  
TAPSE - систолно изместване на трикуспидалния клапен пръстен  
TNF – туморен некротичен фактор

## 1. Въведение

Хроничната обструктивна белодробна болест (ХОББ) е често срещано, предотвратимо и изискващо лечение заболяване, което се характеризира с постоянни респираторни симптоми и ограничение на въздушния поток. Понастоящем ХОББ е третата водеща причина за смърт в световен мащаб с прогноза за увеличаване на тежестта ѝ през следващите десетилетия поради продължаващата експозиция на рискови фактори за ХОББ и застаряването на населението.

Приложението на изкуствен интелект (ИИ) в медицината и в частност - в белодробната медицина, придобива все по-голяма популярност. Машинното обучение (МО) е част от ИИ, при което компютрите използват статистически методи за самообучение без да са изрично програмирани. В областта на респираторната медицина са проведени няколко проучвания с акцент обструктивни заболявания и белодробна фиброза по отношение на диагноза, стадиране, екзацербации и преживяемост.

Извършването на информирана клинична преценка въз основа на съществуващите данни е в основата на медицината, базирана на доказателствата. МО може да бъде следващата стъпка в персонализираната медицина при ХОББ чрез изчисляване на точния риск (например риск от обостряне, смърт) на всеки пациент (въз основа на наличните параметри като белодробна функция, клинични данни, демографски данни, предходни обостряния и т.н.), като по този начин предоставя прогноза/риск на ниво пациент въз основа на индивидуалните характеристики (индивидуален подход).

Настоящият научен труд изследва възможностите на изкуствения интелект и машинното обучение при пациенти с ХОББ върху прогнозирането на клинично значими крайни точки и анализира резултатите спрямо настоящите ръководства. След провеждане на литературен обзор върху рисковите фактори, свързани с влошаване на прогнозата при ХОББ (екзацербации, смърт) е проведена детайлна дескриптивна статистика върху широк набор от параметри, включващ анамнестични данни (демографски показатели), спирометрични показатели, качество на живот, хематологични, ехокардиографски, ехографски и електрокардиографски (ЕКГ) показатели при пациенти с ХОББ преди и след физическо усилие (6-минутен тест с ходене).

Създаването на модели с машинно обучение, използвайки единични характеристики и впоследствие - съвкупност от

параметри за прогнозирането на риска при пациенти с ХОББ, анализ и сравнение на резултатите от създадените модели и спрямо настоящите препоръки, където е възможно, демонстрират възможностите на този вид методология. Не на последно място, проведенят фенотипен анализ посредством три метода (метод на главните елементи, K-means клъстеризация и йерархична клъстеризация) очертава насоките при групиране на пациентите с ХОББ.

## 2. Цел и задачи

### Цел

Целта на настоящия научен труд е да се демонстрират възможностите от приложението на изкуствен интелект и по-специално - машинно обучение, при пациенти с ХОББ върху прогнозирането на клинично значими крайни точки и да се анализират резултатите спрямо настоящите ръководства и препоръки, където това е възможно.

### Задачи

1. Да се проведе **литературен обзор** върху рисковите фактори, свързани с прогнозата при ХОББ (екзацербации, смърт) и приложението на ИИ в белодробната медицина и в частност - при ХОББ;
2. Да се съберат **клинично значими показатели** за оценка на тяхната прогностична стойност чрез провеждане на широк набор от изследвания, включващи анамнестични данни (демографски показатели, честота на предходни екзацербации), спирометрични показатели, качество на живот и симптоми, хематологични, ехокардиографски, ехографски и електрокардиографски (ЕКГ) показатели и функционална оценка (тест с натоварване) при пациенти с ХОББ;
3. Да се проведе детайлна **дескриптивна статистика** на получените параметри в изследваната популацията. Да се анализират възможностите на МО, свързани с отстраняването на проблеми в набора от данни (например липсващи данни);
4. Да се анализират възможностите на машинното обучение, използвайки **единични параметри** върху прогнозирането на риска при пациенти с ХОББ и определяне на **оптимална гранична точка** на индивидуален параметър с цел максимална информативност;
5. Да се демонстрират възможностите на машинното обучение, използвайки **съвкупност от параметри** върху прогнозирането на риска при пациенти с ХОББ;
6. Да се анализират **резултатите** от създадените модели един спрямо друг и спрямо настоящите препоръки, където е възможно;
7. Да се проведе **фенотипен анализ** посредством метод на главните елементи, K-means клъстеризация и йерархична клъстеризация при пациентите с ХОББ.

### 3. Материали и методи

Това е проспективно обсервационно проучване. Включени са пациенти на възраст над 40 години, настоящи или бивши пушачи с повече от 10 пакетогодини тютюнопушене и диагностицирани с ХОББ с постбронходилаторно отношение ФЕО1/ФВК <0,7. Пациентите е трябвало да бъдат в стабилно състояние (без екзацербация) най-малко един месец. Изключващи критерии са наличието на други значими белодробни заболявания (муковисцидоза, тежки бронхоектазии, карцином или рестриктивно белодробно заболяване), системно възпалително заболяване или невъзможност за спазване на процедурите (напр. попълване на въпросниците за качество на живот, провеждане на 6-минутен тест с ходене).

Пациентите бяха проследени в продължение на 12 месеца, по време на които е предоставена информация за броя и тежестта на екзацербациите с помощта на структуриран въпросник. Пациентът е отчетен като изгубен при проследяване, ако не могат да бъдат получени данни за едногодишния период на проспективно наблюдение.

Крайни точки в това проучване са:

За моделите с определяне на оптимална гранична точка (еднофакторните модели):

- Композитна крайна точка, състояща се от времето до първата умерено-тежка или тежка екзацербация или смърт.

За моделите с МО (многофакторните модели):

- Наличие на екзацербация при едногодишно проследяване (да/не);
- Категория на екзацербатор/неекзацербатор по GOLD (т.е. за „екзацербатор“ се приема наличие на тежка екзацербация или  $\geq 2$  умерено-тежки екзацербации при едногодишно проследяване) (да/не);
- Брой екзацербации при едногодишно проследяване (числова стойност);
- Брой екзацербации при едногодишно проследяване (категория);

За моделите с определяне на оптимална гранична точка настоящият научен труд е с фокус върху ехокардиографски и ехографски параметри. По този начин ще бъде демонстрирана оценката на показатели с налична гранична точка (например RARm

≥20 mmHg за установяване на ПХ) и такива без установена гранична крайна точка (например диафрагмална подвижност).

### **Анамнестични данни**

Събрани са следните анамнестични данни:

- демографски данни: пол (мъж/жена); възраст (години); тютюнопушене (бивш пушач/настоящ пушач); пакетогодини тютюнопушене; тегло (kg); ръст (cm); индекс на телесна маса (BMI);
- терапия: употреба на инхалаторни кортикостероиди (ICS); дългодействащи бета-агонисти (LABA); дългодействащи мускаринови антагонисти (LAMA); краткодействащи мускаринови антагонисти (SAMA); инхибитори на фосфодиестераза 4 (PDI4); домашна кислородотерапия; системна употреба на муколитичи; дългосрочна терапия с макролиди; без терапия (при липсата на лечение с изброените терапевтични групи);
- придружаващи заболявания: предходна диагноза на бронхиална астма; АСО; емфизем; бронхиектазии; хроничен бронхит; OSA; артериална хипертония; остеопороза; дислипидемия; хиперурикемия; депресия; друго. Изчислен е коморбидният индекс на Charlson.

Charlson индексът на коморбидност представлява претеглен индекс за прогнозиране на риска от десетгодишна смъртност при пациенти със специфични коморбидни състояния. Индексът включва 19 състояния, на всяко от които е присвоено тегло от 1 до 6 въз основа на прогнозния коефициент на риск за смърт (Таблица 1). Тези тегла се сумират, за да се получи финалният резултат.

**Таблица 1. Charlson индекс на коморбидност (CCI)**

<b>Коморбидност</b>	<b>Точки</b>
Предходен инфаркт на миокарда	1
Застойна сърдечна недостатъчност	1
Периферно съдово заболяване	1
Мозъчно-съдово заболяване	1
Деменция	1
Хронично белодробно заболяване	1
Ревматологично заболяване	1
Пептична язва	1
Лекостепенно чернодробно заболяване	1

Диабет	1
Мозъчно-съдов инцидент (хемиплегия)	2
Умерено до тежко бъбречно заболяване	2
Диабет с хронични усложнения	2
Карцином без метастази	2
Левкемия	2
Лимфом	2
Умерено или тежко чернодробно заболяване	3
Метастатичен солиден тумор	6
Синдром на придобита имунна недостатъчност (СПИН)	6

### Спирометрични показатели

Пациентите бяха инструктирани да се въздържат от краткодействащи  $\beta$ 2-агонисти поне 6 часа, дългодействащи  $\beta$ 2-агонисти поне 12 часа, дългодействащи холинолитици поне 24 часа и краткодействащи холинолитици поне 12 часа преди спирометрията (450). Постбронходилататорната спирометрия е осъществена 15-30 минути след инхалация на 400 mcg Salbutamol в съответствие с препоръките на ERS/ATS (450). Измерени бяха пре- и постбронходилататорните стойности на форсирания витален капацитет (ФВК), форсирания експираторен обем за първата секунда (ФЕО1) и отношението ФЕО1/ФВК. Обструкцията е класифицирана според тежестта на ограничението на въздушния поток съобразно постбронходилататорния ФЕО1 както следва: лека ( $\geq 80\%$  от предвидения); умерена ( $80 > \text{ФЕО1} \geq 50\%$  от предвидения); тежка ( $50\% > \text{ФЕО1} \geq 30\%$  от предвидения); много тежка ( $< 30\%$  от предвидения) (451).

### Екзацербации на ХОББ

Събрана е информация за броя на *тежките* екзацербации (хоспитализации), както и за броя на *умерено-тежките* екзацербации (прием на антибиотично лечение и/или системни кортикостероиди поради влошаване на белодробната симптоматика без хоспитализация) в последната година (98). Общият брой на екзацербациите е изчислен чрез сбор между тежките и умерено-тежките екзацербации в предходната година.

Пациентите са разделени на клинични фенотипи съгласно критериите на GesEPOC (448):

- неекзацербатор: пациенти с 0 или 1 екзацербации през предходната година;

- екзацербатор с емфизем - пациенти с поне две екзацербации през предходната година и клинични/рентгенологични или функционални данни за емфизем;
- екзацербатор с хроничен бронхит - пациенти с поне две екзацербации, с кашлица и експекторация през поне 3 месеца от годината в продължение на две последователни години;
- Астма-ХОББ овърлап - пациенти със съпътстваща диагноза астма и/или еозинофилия в периферна кръв  $\geq 300$  клетки/ $\mu\text{L}$  (449).

### Качество на живот и симптоми

Качеството на живот беше оценено с помощта на mMRC скалата и CAT (COPD Assessment Test) въпросника. Преди попълването им беше отговорено на всички възникнали въпроси, като беше обяснено, че няма правилни и грешни отговори. Пациентите бяха класифицирани според GOLD като такива с по-малко симптоми (CAT <10) и задух (mMRC 0-1) и с повече симптоми (CAT  $\geq 10$ ) и задух (mMRC  $\geq 2$ ).

### mMRC скала на задуха

Участниците в проучването бяха помолени да попълнят mMRC скала на задуха (Таблица 2) като отбележат при какви усилия получават задух. Резултат от „0“ и „1“ съответства на ниска степен на задух, а резултат  $\geq 2$  класифицира пациенти с висока степен на задух.

**Таблица 2. mMRC скала на задуха**

<b>mMRC скала на задуха (диспнея)</b>	
Моля, отбележете степента, която най-много съответства на Вашето състояние:	
<b>Тежест</b>	<b>Степен на задуха според физическата активност</b>
0	Нямам задух освен при тежко физическо натоварване.
1	Задух при бързо ходене или при леко изкачване.
2	Ходя по-бавно от връстниците си на равен терен поради задух или трябва да спирам поради задух при ходене с присъщата ми бързина.
3	Спирам поради задух след около 100 м или след няколко минути при ходене на равни.
4	Не мога да напускам дома поради задух или задух при обличане/събличане.

### CAT (COPD Assessment Test) въпросник

CAT въпросникът съдържа 8 въпроса, на които пациентът отговаря сам (Фигура 1). Въпросите са под формата на две противоположни твърдения с 6-степенна скала между тях (от 0 до 5). Всяка степен отговаря на една точка. Общият брой точки се сумира. Максималният брой точки е 40. Резултат <10 съответства на пациенти с малко симптоми, а резултат  $\geq 10$  класифицира високосимптомни пациенти.

CAT мест

Вашето име:  Днешна дата:  

**Какво е състоянието на Вашата ХОББ (хронична обструктивна белодробна болест)? Направете теста за оценка на ХОББ (COPD Assessment Test™, CAT)**

Този въпросник ще помогне на Вас и Вашия здравен специалист да измерите влиянието, което ХОББ (хронична обструктивна белодробна болест) оказва върху Вашето здраве и ежедневие. Вашите отговори и точките от теста могат да се използват от Вас и Вашия здравен специалист да помогнат за подобряване контрола върху Вашата ХОББ и получаване на най-голяма полза от лечението.

За всяка от точките по-долу поставете знак X в кутийката, която най-добре описва състоянието Ви в момента. Не забравяйте да изберете само по един отговор на всеки въпрос.

Пример: много съм щастлив(а)  0  1  2  3  4  5 Много съм тъжен(на)

Никога не кашлям	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5	Кашлям непрекъснато	<input type="radio"/>
Изобщо нямам хранения в гърдите си	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5	Целите ми гърди са пълни с хранения	<input type="radio"/>
Изобщо не чувствам стягане в гърдите	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5	Чувствам голямо стягане в гърдите	<input type="radio"/>
Когато вървя по нагорнище или изкачвам един етаж стълби, не се задъхвам	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5	Когато вървя по нагорнище или изкачвам един етаж стълби, се задъхвам много	<input type="radio"/>
Не съм ограничен(а) в извършването на каквито и било дейности у дома	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5	Много съм ограничен(а) в извършването на дейности у дома	<input type="radio"/>
Уверен(а) съм, когато излизам от дома си, въпреки състоянието на белите ми дробове	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5	Никак не съм уверен(а), когато излизам от дома си, поради състоянието на белите ми дробове	<input type="radio"/>
Спя спокойно	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5	Не спя спокойно поради състоянието на белите ми дробове	<input type="radio"/>
Имам много енергия	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5	Изобщо нямам енергия	<input type="radio"/>
			<input type="text"/>

COPD Assessment Test™ и логото CAT са запазена марка на GlaxoSmithKline © GlaxoSmithKline. Всички права запазени.

ОБЩ БРОЙ ТОЧКИ

## Фигура 1. CAT въпросник

### Симптоми

Наличието на симптоми беше оценено с помощта на бинарна променлива за всеки от четирите кардинални симптома – кашлица, свирене в гърдите, експекторация и задух. За всеки от симптомите пациентите са попитани дали усещат неговото присъствие в ежедневието си. В случай на положителен отговор, симптомът е отбелязан като присъстващ.

### **Хематологични показатели**

В деня на началното посещение е взета венозна кръв. Изследвани са следните параметри:

- Пълна кръвна картина (ПКК) - еритроцити ( $\times 10^{12}/L$ ); левкоцити ( $\times 10^9/L$ ); хемоглобин (g/l); хематокрит (%); неутрофили (абсолютен брой и %); лимфоцити (абсолютен брой и %); базофили (абсолютен брой и %); еозинофили (абсолютен брой и %); моноцити (абсолютен брой и %); тромбоцити ( $\times 10^9/L$ );
- Биохимия и хемостаза – фибриноген (g/L), CRP (mg/dL),  $\alpha 1$ -антитрипсин (g/L).

Допълнително бяха изчислени следните отношения:

- съотношение неутрофили/лимфоцити (neutrophil/lymphocyte ratio, NLR) като отношението между абсолютния брой неутрофили и лимфоцити;
- съотношение тромбоцити/лимфоцити (platelet/lymphocyte ratio, PLR) като отношението между тромбоцитите и абсолютния брой лимфоцити;
- съотношение еозинофили/базофили (eosinophil/basophil ratio, EBR) като отношението между абсолютния брой еозинофили и базофили;
- съотношение еозинофили/лимфоцити (eosinophil/lymphocyte ratio, ELR) като отношението между абсолютния брой еозинофили и лимфоцити.

### **Ехокардиографски показатели**

При всички пациенти беше проведена двуизмерна трансторакална доплерова ехокардиография в покой от опитен кардиолог. Използван е секторен трансдюсер с честота 2–3.5 MHz в 2D и M-Mode режими. Функцията на лявата камера е оценена чрез следните параметри:

- морфология – теледиастолен размер (ТДР); телесистолен размер (ТСП); дебелина на септум; задна стена на лява камера (ЗСЛК); размер и площ на ляво предсърдие;

- налягане в ляво предсърдие (LAP) – разликата в систолните налягания между лява камера и ляво предсърдие на нивото на максималната MR скорост;

- фракция на изтласкване (EF, ФИ) – изтласкваният обем от края на диастолата при всяка контракция, изчислен по метода на Simpson:

$$EF = (EDV-ESV)/EDV \times 100$$

където EF е фракция на изтласкване, EDV - краен диастолен обем и ESV – краен систолен обем. Левокамерна фракция на изтласкване под 55% е отчетена като систолна дисфункция на лявата камера (LVSD).

- фракция на съкращение (ФС) - процентната промяна в размера на лявата камера при всяка контракция, изчислена по формулата:

$$(TDP - TCP / TDP) \times 100$$

Стойности <28% са класифицирани като намалена ФС.

- E/A - диастолично пълнене на лявата камера, отношение между пиковата скорост на митралния поток на ранната вълна на бързо пълнене (E) и пиковата скорост на късната вълна на пълнене, причинена от предсърдната контракция (A);

- децелерационно време (DT) - времевият интервал от пика на E-вълната до нейната прогнозна базова линия.

Функцията на дясната камера е оценена чрез следните параметри:

- морфология - теледиастолен размер на дясна камера, свободна стена на дясна камера;
- систолно изместване на трикуспидалния клапен пръстен (TAPSE) – измерено в M-mode при апикален четирикухинен срез като систолното лонгитудинално изместване на латералния трикуспидален пръстен към върха;
- време на акцелерация (msec);
- време на изтласкване (msec);
- средно налягане в пулмоналната артерия (PAPm) ;
- диаметър на дясна пулмонална артерия (mm);
- долна празна вена – минимален и максимален диаметър (mm).

Непосредствено след провеждане на 6MWT са оценени средното налягане в пулмоналната артерия, TAPSE, време на акцелерация, време на изтласкване.

## **Ехографски показатели**

За изследване на инспираторната подвижност на диафрагмата е използван конвексен трансдюсер с честота 3.5-5-7.5 МHz, през трансхепатален прозорец в дясно и транслиенален в ляво. Изследвана е максималната подвижност на лява и дясна диафрагма в покой и след 6MWT.

Директното измерване на диафрагмалната подвижност се извършва с помощта на B-mode. Разположението на трансдюсера в долните междуребрива между предна и задна аксиларна линия позволява перпендикулярна на движението на диафрагмата посока на ултразвуковия лъч и свежда до минимум пречката от чревни газове при изследване на лявата хемидиафрагма. Най-високата точка на купола се маркира на най-високото място по време на максимално издишване и на най-ниското място по време на максимално вдишване (452).

## **Електрокардиографски показатели**

Проведена е стандартна ЕКГ с 12 отвеждания в покой със скорост 25 mm/s. Изследвани са следните параметри: сърдечна честота в покой; височина на р-вълната (в mm), други отклонения (ритъмни и проводни нарушения, левокамерна/деснокамерна хипертрофия).

За левокамерна хипертрофия е използван критерият на Соколов-Лион (Sokolow-Lyon): височината на S във V1 + височината на R във V5 или V6 >3.5 mV (35 mm). Деснокамерна хипертрофия е приета при наличие на дясна ос на QRS комплекса и отношение на височината на R във V1 към височината на S във V1 >1.

## **Оценка на физическия капацитет**

За оценка на физическия капацитет е проведен 6MWT според стандартизираните международни указания (400). Използван е прав, затворен коридор с дължина 30 m, маркиран през 3 m. Пациентите са били в покой поне 10 минути преди началото на теста и са получили стандартни указания за теста. Беше отговорено на всички възникнали въпроси. Записани са следните показатели:

- продължителност на 6MWT (min);
- изминато разстояние (m);

- кислородна сатурация (%) в началото и в края на теста. Изчислена е десатурацията като разлика между кислородната сатурация в края спрямо началото на теста.

След провеждане на 6MWT е изчислен BODE индексът. BODE индексът е многомерна система за класификация, използвана в медицината за оценка на тежестта и прогнозата на ХОББ. Той включва индекс на телесната маса (B, BMI), обструкция на въздушния поток (O, измерена чрез ФЕО1), диспнея по mMRC скалата (D) и физическия капацитет (E, измерен чрез 6MWT). Индексът обединява тези четири параметъра в една оценка, която варира от 0 до 10, като резултатът корелира с тежестта.

BMI  $\leq 21$  kg/m<sup>2</sup> добавя 1 точка. Наличието на тежка обструкция (ФЕО1 50-64% - 1 точка, 36-49% - 2 точки,  $\leq 35\%$  - 3 точки), по-високият mMRC резултат (категория „2“ – 1 точка, категория „3“ – 2 точки, категория „4“ – 3 точки) и намаленото изминато разстояние на 6MWT (250-349 m – 1 точка, 150-249 m – 2 точки,  $\leq 149$  m – 3 точки) също добавят от 1 до 3 точки.

## **Проследяване**

Пациентите са проследени в продължение на 12 месеца, като на 6-ти и 12-ти месец е събрана информация за броя и тежестта на екзацербациите през съответния период със структуриран въпросник.

## **Статистическа обработка**

### *Обща информация*

Настоящият научен труд изследва възможностите на машинното обучение за прогнозиране на честотата на екзацербациите и смъртността в три направления:

- създаване на еднокомпонентни модели – модели с един предиктор. Изследвани са възможностите за подобряване на точността с помощта на определяне на оптимални гранични точки;
- създаване на многокомпонентни модели – модели с повече от един предиктор. Това са традиционно използваните модели за прогнозиране;
- клъстерен анализ – създаване на групи от пациенти с общи характеристики, които могат да имат различно протичане на болестта (екзацербации).

## Софтуерен пакет

Всички статистически анализи са проведени с помощта на RStudio 4.2.2.

## Обработка на липсващи данни

За всеки показател е оценено наличието на липсващи стойности. При наличие на повече от 50% липсващи стойности за един показател, целият показател се премахва от последващи анализи. При липсващи данни до 50% променливата се тества за нормалност чрез теста на Shapiro-Wilk. При наличие на нормално разпределение, на липсващите данните се приписва средната стойност на променливата, а при липса на нормално разпределение – медианата.

## Определяне на гранични точки (cut-off)

Изборът на оптимални гранични точки за отделните параметри е извършен с помощта на пакета *cutpointr* (<https://cran.r-project.org/web/packages/cutpointr/vignettes/cutpointr.html>) с цел максимизиране на метричната функция *Youden's J* (*method = maximize\_metric*, *metric = youden*). Последната се изчислява по следната формула:

$$J = \text{сензитивност} + \text{специфичност} - 1$$

и се използва при определяне на ROC крива. Изчислената гранична стойност е използвана за разделяне на групите и извършване на последващ статистически анализ.

## Оценяване на предиктивните модели за машинно обучение

За оценка на създадените предиктивни модели са използвани следните показатели:

- за категорийни променливи – честота на правилно класифицираните наблюдения, измерена като брой правилно класифицирани наблюдения спрямо общ брой наблюдения;
- за числови променливи – корен от средната квадратична грешка (RMSE). За всяка точка с данни формулата за RMSE изчислява повдигнатата на квадрат разлика между действителната стойност на точката и стойността на точката на кривата с най-доброто прилягане. След това сумата от остатъците се разделя на общия брой точки и от този резултат се изчислява квадратния корен.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

### *Възпроизводимост на резултатите*

Случайните числа, генерирани от компютър, всъщност не са истински случайни. Те се получават чрез сложни математически алгоритми, наречени генератори на случайни числа (Random Number Generators или RNG) и могат да бъдат използвани в различни статистически и аналитични процеси. Тези алгоритми започват да работят с определена стойност, наречена „seed“ или начално състояние.

“Seed” представлява инструмент, който помага за управление и контрол на случайността при изчисленията, като същевременно гарантира възпроизводимостта и сравнимостта на резултатите. Когато стойността на „seed“ е фиксирана, например с определено число или текст, генераторът на случайни числа ще създава всички последващи случайни числа, използвайки този конкретен „seed“. Това означава, че при еднакъв „seed“ се получават същите случайни числа всеки път, когато генераторът на случайни числа се стартира.

С цел възпроизвеждане на резултатите ще бъде използван *seed* за случайните числа (*set.seed(100)*).

### *Разделяне на извадката*

Извадката е разделена на тренировъчен набор от данни, който е използван за обучение на моделите, и тестови набор от данни, който е използван за оценка на създадените модели. Извадката е разделена в отношение 70:30% с помощта на *seed* за случайните числа, т.е. тренировъчните и тестовите данни за всички модели са еднакви.

### **Описание на алгоритмите за машинно обучение**

Същността на машинното обучение се крие в способността му да извлича ценна информация и да генерира предиктивни модели от големи и сложни масиви от данни, придавайки стойност на необработената информация. Настоящият научен труд използва както обучение с надзирател (supervised) за създаване на предиктивни модели, така и обучение без надзирател (unsupervised)

за провеждане на фенотипен анализ. В този раздел са представени ключовите концепции, които са в основата на използваните алгоритми.

Използвани са следните алгоритми за машинно обучение:

- за регресия – линейна регресия, дърво на решението (decision tree), Random Forest, невронни мрежи, XGBoost и CatBoost;
- за класификация – логистична регресия, метод на k-най-близките съседи (KNN), дърво на решението, Random Forest, Naïve Bayes, невронни мрежи, поддържаща векторна машина (SVM), XGBoost и CatBoost.

В настоящия труд са демонстрирани и двата основни вида неконтролирано обучение – метод на главните елементи (principal components analysis, PCA), при който данните се трансформират, като се запазва възможно най-много от оригиналната информация, и клъстеризация (clustering) – широк клас от методи за откриване на подгрупи. Най-честите техники за клъстеризация включват метод на K-средните и йерархична клъстеризация.

## 4. Резултати

Резултатите в настоящия научен труд са структурирани по следния начин:

- дескриптивна статистика - описание на променливите (включва описание на характеристиките на пациентите и заболяването, заедно с начина им на кодиране и липсващи стойности) и последващ еднокомпонентен анализ (включва изследване на корелации между отделните променливи, определяне на оптимална гранична стойност и прогностична стойност на отделна променлива при прогнозиране на честотата на екзацербациите и композитната крайна точка), разделени по групи:
  - демографски показатели;
  - честота на предходни екзацербации;
  - спиromетрични показатели;
  - качество на живот и симптоми;
  - хематологични показатели;
  - ехокардиографски показатели;
  - ехографски показатели;
  - ЕКГ показатели;
  - оценка на физическия капацитет.
- многокомпонентен анализ – създаване на модели за машинно обучение за прогнозиране на:
  - наличие на екзацербация („да“ срещу „не“);
  - риск по GOLD (група „Е“ срещу групи „А“ и „В“)
  - честота на екзацербации (числова променлива)
  - честота на екзацербации (категорийна променлива)
- фенотипизиране – включва резултати от метода на главните елементи, метода на К-средните и йерархична клъстеризация.

### 4.1. Дескриптивна статистика

#### Анамнестични показатели

На Таблица 3 са представени използваните променливи, техният вид, начинът на кодиране и мерните единици.

**Таблица 3. Характеристики на променливите (демографски показатели)**

Променлива	Вид	Начин на кодиране	Мерни единици	Средна стойнос
------------	-----	-------------------	---------------	----------------

				<b>T</b>
Пол	Категорийна	„0“ – жена „1“ – мъж	-	0.625 (62.5% мъже)
Възраст	Числова	Закръгляване до цяло число	Години	66.6
Тютюнопушене	Категорийна	„0“ – бивш пушач „1“ – настоящ пушач	-	0.344 (34.4% настоящи пушачи)
Пушачески стаж	Числова	Закръгляване до цяло число	Пакетогодини = брой цигари на ден * брой години тютюнопушене	24.38
Тегло	Числова	Закръгляване до цяло число	Килограм	72.41
Ръст	Числова	Закръгляване до цяло число	Сантиметър	167.1
ВМІ	Числова	Закръгляване до втори знак след десетичната запетая	-	25.77
Година на диагностициране на ХОББ	Числова	Цяло число	Година	2014.5
<b>Придружаващи заболявания</b>				
Индекс на коморбидност на Charlson	Числова	Цяло число	-	3.85
Бронхиална астма	Категорийна	„0“ – не „1“ – да	-	0.115 (11.5%)
Астма-ХОББ припокриване (АСО)	Категорийна	„0“ – не „1“ – да	-	0.125 (12.5%)
Емфизем	Категорийна	„0“ – не „1“ – да	-	0.573 (57.3%)

Бронхиектазии	Категорийна	„0” – не „1“ – да	-	0.063 (6.3%)
Хроничен bronхит	Категорийна	„0” – не „1“ – да	-	0.635 (63.5%)
Обструктивна сънна апнея	Категорийна	„0” – не „1“ – да	-	0.063 (6.3%)
Артериална хипертония	Категорийна	„0” – не „1“ – да	-	(0.74) 74.0%
Остеопороза	Категорийна	„0” – не „1“ – да	-	0.021 (2.1%)
Дислипидемия	Категорийна	„0” – не „1“ – да	-	0.177 (17.7%)
Хиперурикемия	Категорийна	„0” – не „1“ – да	-	0.052 (5.2%)
Терапия				
Инхалаторни кортикостероиди	Категорийна	„0” – не „1“ – да	-	0.573 (57.3%)
Дългодействащи бета-агонисти	Категорийна	„0” – не „1“ – да	-	0.823 (82.3%)
Дългодействащи мускаринови антагонисти	Категорийна	„0” – не „1“ – да	-	0.802 (80.2%)
Краткодействащи и мускаринови антагонисти	Категорийна	„0” – не „1“ – да	-	0
Инхибитори на фосфодиестераза 4	Категорийна	„0” – не „1“ – да	-	0.01 (1%)
Домашна кислородотерапия	Категорийна	„0” – не „1“ – да	-	0.063 (6.3%)
Системна употреба на муколитици	Категорийна	„0” – не „1“ – да	-	0
Дългосрочна терапия с макролиди	Категорийна	„0” – не „1“ – да	-	0
Без терапия	Категорийна	„0” – не „1“ – да	-	0.01 (1%)

Променливите, свързани с лечението „Краткодействащи мускаринови антагонисти“, „Инхибитори на фосфодиестераза 4“,

„Системна употреба на муколитици“, „Дългосрочна терапия с макролиди“ и „Без терапия“ са изключени от анализа поради нулева или близка до нулата вариация на предиктора, т.е. липса на информативност.

В тази група няма липсващи стойности и допълнителна обработка на данните не се налага.

Най-значими корелации се наблюдават между възрастта и индекса на коморбидност на Charlson ( $r = 0.67$ ), между възрастта и броя на пакетогодините тютюнопушене ( $r = 0.42$ ) и броят на пакетогодините тютюнопушене и BODE индекса ( $r = 0.35$ ). BMI корелира най-силно с абсолютния лимфоцитен брой ( $r = 0.25$ ) и хематокрита ( $r = 0.25$ ), докато индекса на коморбидност на Charlson освен с възрастта, показва връзка с изминатото разстояние при 6MWT ( $r = -0.36$ ) и САТ ( $r = 0.31$ ). Продължителността на ХОББ (времето от поставянето на диагнозата) показва връзка със скалата на задуха ( $r = 0.31$ ), BODE индекса ( $r = 0.28$ ) и САТ ( $r = 0.22$ ).

По отношение на придружаващите заболявания с най-висока честота е артериалната хипертония (74.0%). Средната стойност на Charlson индекса на коморбидност е 3.85, но трябва да се уточни, че някои от показателите в него (например туморни заболявания) са изключващ критерии. Тъй като наличието на ХОББ носи „1“ точка, минималният възможен резултат е „1“. На този фон едва 3.1% (3/96) имат „1“ точка, т.е. не отговарят на никой от другите критерии.

### Честота на предходни екзацербации

На Таблица 4 са представени използваните променливи, техният вид, начинът на кодиране и мерните единици.

**Таблица 4. Характеристики на променливите (предходни екзацербации)**

Променлива	Вид	Начин на кодиране	Мерни единици	Средна стойност
Умерено-тежки екзацербации	Числова	Цяло число	брой	0.70
Тежки екзацербации	Числова	Цяло число	брой	0.98
Общ брой екзацербации	Числова	Цяло число	брой	1.68
Фенотипи	Категорийна	„0“ –	-	-

		неекзацербатор „1“ – АСО „2“ – екзацербатор с емфизем „3“ – екзацербатор с хроничен бронхит		
--	--	---	--	--

В тази група няма липсващи стойности и допълнителна обработка на данните не се налага.

Най-значими корелации честотата на умерено-тежките екзацербации показва с САТ ( $r = 0.45$ ), mMRC ( $r = 0.43$ ) и ФЕО1 ( $r = -0.37$ ), докато честотата на тежките екзацербации корелира значимо със същите параметри, но в по-ниска степен:  $r = 0.21$ ,  $r = 0.29$  с  $r = -0.22$  съответно за САТ, mMRC и ФЕО1. По отношение на едногодишното проследяване честотата на тежките екзацербации в предходната година е показателят с най-висока корелация с общия брой екзацербации ( $r = 0.52$ ), докато умерено-тежките екзацербации имат коефициент на корелация  $r = 0.34$ .

Само 16.7% (16/96) от пациентите нямат екзацербации в предходната година. Разпределението по GOLD е представено на Таблица 5. Повечето участници (73.9%) имат висок риск от екзацербация.

**Таблица 5. Разпределение на участниците според риска от екзацербации по GOLD**

Група Е 73.9% (71/96)	
Група А 11.5% 11.5% (11/96)	Група В 14.6% (14/96)

### Спирометрични показатели

На Таблица 6 са представени използваните променливи, техният вид, начинът на кодиране и мерните единици. На Таблица 7 са представени липсващите стойности и последващото действие за тяхната обработка.

**Таблица 6. Характеристики на променливите (спирометрични показатели)**

Променлива	Вид	Начин на кодиране	Мерни единици	Средна стойност
ФЕО1 преди бронходилатация	Числова	Цяло число	%	51.3
ФЕО1 след бронходилатация	Числова	Цяло число	%	55.8
ФЕО1/ФВК след бронходилатация	Числова	Закръгляване до втори знак след десетичната запетая	-	0.56
ФВК след бронходилатация	Числова	Цяло число	%	77.9

**Таблица 7. Липсващи стойности (спирометрични показатели)**

Променлива	Липсващи стойности	Действие
ФЕО1 преди бронходилатация	0	-
ФЕО1 след бронходилатация	0	-
ФЕО1/ФВК след бронходилатация	0	-
ФВК след бронходилатация	0	-

В тази група няма липсващи стойности и допълнителна обработка на данните не се налага.

Средната стойност на ФЕО1 е  $55.8 \pm 18.3\%$ . 7.3% от участниците (7/96) имат постбронходилататорен ФЕО1 <30%, 33.3% (32/96) – между 30 и 50%, 49.0% (47/96) – между 50 и 80% и 10.4% (10/96) – над 80%. Средната стойност на ФВК е  $77.9 \pm 22.5\%$ . Средното подобрение на ФЕО1 след бронходилатация е 4.4%. Само 13.5% (13/96) от пациентите демонстрират подобрение на ФЕО1  $\geq 10\%$ , а 8.3% (8/96) -  $\geq 12\%$ .

Белодробната функция показва значими корелации с много характеристики на заболяването, които ще бъдат разгледани в съответните раздели. Все пак най-значими корелации ФЕО1 след бронходилатация показва с BODE индекса ( $r = -0.81$ ), mMRC ( $r = -0.68$ ) и CAT ( $r = -0.59$ ).

### **Качество на живот и симптоми**

На Таблица 8 са представени използваните променливи, техният вид, начинът на кодиране и мерните единици.

**Таблица 8. Характеристики на променливите (показатели за качество на живот и симптоми)**

Променлива	Вид	Начин на кодиране	Мерни единици	Средна стойност
mMRC	Категорийна	Цяло число	Точки	-
CAT	Числова	Цяло число	Точки	15.6
Симптом: кашлица	Категорийна	„0” – не „1” – да	-	0.938 (93.8%)
Симптом: свирене в гърдите	Категорийна	„0” – не „1” – да	-	0.719 (71.9%)
Симптом: експекторация	Категорийна	„0” – не „1” – да	-	0.719 (71.9%)
Симптом: задух	Категорийна	„0” – не „1” – да	-	0.823 (82.3%)

В тази група няма липсващи стойности и допълнителна обработка на данните не се налага.

Средната стойност на резултата от CAT теста е  $15.6 \pm 9.2$  точки. По отношение на резултатите от скалата на задуха 11.5% (11/96) отбелязват резултат „0“, 15.6% (15/96) - резултат „1“, 29.2% (28/96) – резултат „2“, 42.7% (41/96) – резултат „3“ и 1.0% (1/96) – резултат „4“.

Аналогично на белодробната функция, качеството на живот също показва значими корелации с много характеристики на заболяването, които ще бъдат разгледани в съответните раздели. Все пак най-значими корелации mMRC показва с BODE индекса ( $r = 0.86$ ), изминатото разстояние при 6MWT ( $r = -0.71$ ), ФЕО1 след бронходилатация ( $r = -0.68$ ) и CAT ( $r = -0.59$ ). Аналогични резултати се наблюдават и за резултатите от CAT –  $r = 0.77$ ,  $r = -0.7$ ,  $r = -0.59$  съответно за BODE индекса, изминатото разстояние при 6MWT и ФЕО1 след бронходилатация.

По отношение на докладваните симптоми повечето пациенти са симптоматични, като с най-голяма честота на докладване е наличието на кашлица (90/96, 93.8%). Най-значими корелации се наблюдават с белодробната функция ( $r = -0.51$  между свирене в гърдите и ФЕО1), физическия капацитет ( $r = -0.46$  между задух и изминатото разстояние на 6MWT) и диафрагмалната подвижност ( $r = -0.39$  за свирене в гърдите).

## Хематологични показатели

На Таблица 9 са представени използваните променливи, техният вид, начинът на кодиране и мерните единици.

**Таблица 9. Характеристики на променливите (хематологични показатели)**

Променлива	Вид	Начин на кодиране	Мерни единици	Средна стойност
Еритроцити	Числова	Закръгляване до първия знак след десетичната запетая	$\times 10^{12}/L$	$4.83 \pm 0.5$
Левкоцити	Числова	Закръгляване до първия знак след десетичната запетая	$\times 10^9/L$	$8.61 \pm 2.32$
Хемоглобин	Числова	Цяло число	g/l	$142.5 \pm 15.1$
Хематокрит	Числова	Закръгляване до втория знак след десетичната запетая	%	$0.45 \pm 0.05$
Неутрофили	Числова	Закръгляване до първия знак след десетичната запетая	$\times 10^9/L$	$5.48 \pm 1.84$
Неутрофили	Числова	Цяло число	%	$63.1 \pm 9.3$
Лимфоцити	Числова	Закръгляване до първия знак след десетичната запетая	$\times 10^9/L$	$2.1 \pm 0.78$
Лимфоцити	Числова	Цяло число	%	$25.1 \pm 7.4$
Базофили	Числова	Закръгляване до втория знак след десетичната запетая	$\times 10^9/L$	$0.05 \pm 0.04$
Базофили	Числова	Цяло число	%	$0.69 \pm 0.49$
Еозинофили	Числова	Закръгляване до първия знак след десетичната запетая	$\times 10^9/L$	$0.23 \pm 0.2$
Еозинофили	Числова	Закръгляване до първия знак след десетичната запетая	%	$2.7 \pm 2.4$
Моноцити	Числова	Закръгляване до първия знак след десетичната запетая	$\times 10^9/L$	$0.73 \pm 0.24$
Моноцити	Числова	Закръгляване до първия знак след десетичната запетая	%	$8.7 \pm 2.5$

Тромбоцити	Числова	Цяло число	$\times 10^6/L$	$269.7 \pm 67.5$
Фибриноген	Числова	Закръгляване до първия знак след десетичната запетая	g/L	$4.02 \pm 0.98$
CRP	Числова	Закръгляване до първия знак след десетичната запетая	mg/L	$4.4 \pm 4.2$
$\alpha 1$ -антитрипсин	Числова	Закръгляване до втория знак след десетичната запетая	g/L	$1.44 \pm 0.25$

В тази група няма липсващи стойности и допълнителна обработка на данните не се налага.

Хемоглобин  $>170$  g/L имат 3 (3.1%) пациента, а  $<120$  g/L – 7 (7.3%) пациента. Няма пациенти с хемоглобин  $>180$  g/l. Левкоцитоза (абсолютен брой левкоцити  $>10.5 \times 10^9/L$ ) имат 13 (13.5%) пациента. Няма пациенти с левкоцитен брой  $<3.5 \times 10^9/L$ . Еозинофилен брой  $>300 \times 10^9/L$  имат 21 (21.9%) пациента, а  $>500 \times 10^9/L$  – 6 (6.3%) пациента.

Фибриноген  $>4$  g/l имат 41 (42.7%) от пациентите, а  $>7$  g/L – 2 (2.1%) пациента. CRP  $>5$  mg/L имат 33 (34.4%) от пациентите, а над 5.5 mg/L – 29 (30.2%). Ниво на  $\alpha 1$ -антитрипсин  $<1.2$  g/l имат 13 (13.5%) от пациентите, а  $<1.0$  g/l – 3 (3.1%) пациенти.

Най-значими корелации в тази група показатели се наблюдава между еозинофиите (%) и кислородната сатурация в покой ( $r = 0.35$ ), между хематокрита и индекса на коморбидност на Charlson ( $r = -0.29$ ) и между лимфоцитите (%) и подвижността на лявата диафрагма в покой ( $r = 0.27$ ).

### Ехокардиографски показатели

На Таблица 10 са представени използваните променливи, техният вид, начинът на кодиране и мерните единици. На Таблица 11 са представени липсващите стойности и последващото действие за тяхната обработка.

**Таблица 10. Характеристики на променливите (ехокардиографски показатели)**

Променлива	Вид	Начин на кодиране	Мерни единици	Средна стойност
Теледиастолен	Числова	Закръгляване	mm	45.9

размер		до първия знак след десетичната запетая		
Телесистолен размер	Числова	Закръгляване до първия знак след десетичната запетая	mm	30.7
Фракция на изтласкване	Числова	Цяло число	%	60.9
Фракция на съкращение	Числова	Цяло число	%	32.4
Дебелина на септума	Числова	Закръгляване до първия знак след десетичната запетая	mm	11.4
Задна стена на лява камера	Числова	Закръгляване до първия знак след десетичната запетая	mm	11.3
Скорост на Е вълната	Числова	Закръгляване до третия знак след десетичната запетая	m/s	0.57
Децелерационно време	Числова	Цяло число	msec	218
Скорост на А вълната	Числова	Закръгляване до третия знак след десетичната запетая	m/s	0.73
Ляво предсърдие	Числова	Закръгляване до първия знак след десетичната запетая	mm	35.2
Площ на ляво предсърдие	Числова	Закръгляване до първия знак след	cm <sup>2</sup>	18.0

		десетичната запетая		
Налягане в ляво предсърдие	Числова	Цяло число	mmHg	11.6
Теледиастолен размер на дясна камера	Числова	Закръгляване до първия знак след десетичната запетая	mm	29.7
Свободна стена на дясна камера	Числова	Закръгляване до първия знак след десетичната запетая	mm	5.6
TAPSE	Числова	Закръгляване до първия знак след десетичната запетая	mm	19.9
Време на акцелерация	Числова	Цяло число	msec	118.6
Време на изтласкване	Числова	Цяло число	msec	282.6
PAPm	Числова	Цяло число	mmHg	22.4
Дясна пулмонална артерия	Числова	Закръгляване до първия знак след десетичната запетая	mm	24.6
Долна празна вена, максимален размер	Числова	Закръгляване до първия знак след десетичната запетая	mm	15.5
Долна празна вена, минимален размер	Числова	Закръгляване до първия знак след десетичната запетая	mm	7.9
TAPSE след 6MWT	Числова	Закръгляване до първия знак след десетичната	mm	21.7

		запетая		
Време на акцелерация след 6MWT	Числова	Цяло число	msec	99.3
Време на изтласкване след 6MWT	Числова	Цяло число	msec	258.9
PAPm след 6MWT	Числова	Цяло число	mmHg	26.2
Наличие на пулмонална хипертония	Категорийна	„0” – не „1“ – да	-	0.521 (52.1%)

**Таблица 11. Липсващи стойности (ехокардиографски показатели)**

Променлива	Липсващи стойности	Действие
Теледиастолен размер	1	медiana (45.0)**
Телесистолен размер	1	медiana (30.0)**
Фракция на изтласкване	0	-
Фракция на съкращение	0	-
Дебелина на септума	0	-
Задна стена на лява камера	0	-
Скорост на E вълната	2	медiana (0.55)**
Децелерационно време	0	-
Скорост на A вълната	5	средна стойност (0.73)*
Ляво предсърдие, mm	0	-
Площ на ляво предсърдие	7	медiana (17.0)**
Налягане в ляво предсърдие	0	-
Теледиастолен размер на дясна камера	3	медiana (30.0)**
Свободна стена на дясна камера	6	медiana (5.0)**
TAPSE	1	медiana (20.0)**
Време на акцелерация	2	медiana (115.0)**
Време на изтласкване	2	средна стойност (282.6)*
PAPm	0	-
Дясна пулмонална артерия	11	средна стойност (24.6)*
Долна празна вена,	2	медiana (15.0)**

максимален размер		
Долна празна вена, минимален размер	2	медиана (7.75)**
TAPSE след 6MWT	11	медиана (22)**
Време на акцелерация след 6MWT	1	медиана (100)**
Време на изтласкване след 6MWT	1	медиана (260)**
PAPm след 6MWT	0	-
Наличие на пулмонална хипертония	0	-

\* - наличие на нормално разпределение - на липсващите данните се приписва средната стойност на променливата

\*\* - липса на нормално разпределение – на липсващите данните се приписва медианата на променливата

### Ехокардиографски параметри на лява камера и ляво предсърдие

На Таблица 12 са представени основните ехокардиографски параметри на лява камера и ляво предсърдие. Намалена фракция на изтласкване (<55%) имат 16.7% (16/96), а намалена фракция на съкращение (<28%) – 14.6% (14/96) от пациентите.

### **Таблица 12. Основни ехокардиографски параметри на лява камера и ляво предсърдие**

Характеристика	Стойност ( $\pm$ SD)
ТДР	45.9 $\pm$ 6.2 mm
ТСР	30.7 $\pm$ 5.8 mm
ФИ	60.9 $\pm$ 7%
ФС	32.4 $\pm$ 4.9%
Септум	11.4 $\pm$ 1.4 mm
Задна стена на лява камера	11.3 $\pm$ 1.3 mm
Площ на ляво предсърдие	18.0 $\pm$ 4 mm <sup>2</sup>
Размер на ляво предсърдие	35.2 $\pm$ 3.8 mm
Е/А	0.80 $\pm$ 0.20
DT	218 $\pm$ 42 msec
LAP	11.6 $\pm$ 1.9 mmHg

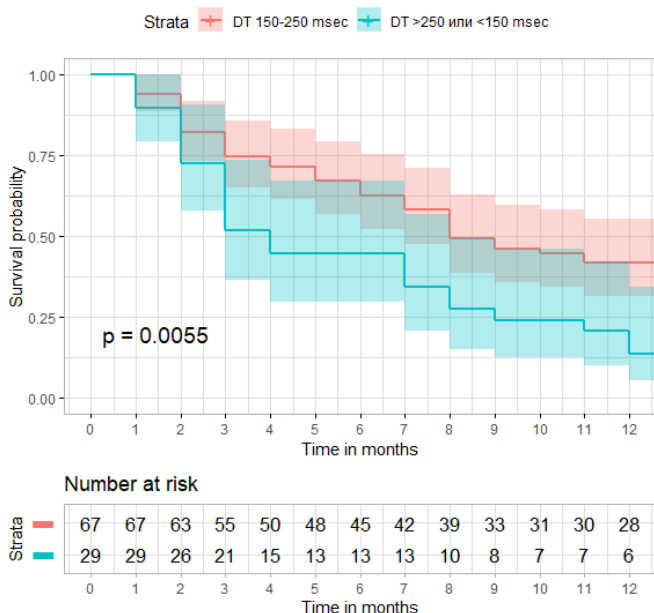
Съкращения: DT – децелерационно време, Е/А - диастолично пълнене на лявата камера, LAP - налягане в ляво предсърдие, ТДР - теледиастолен размер, ТСР - телесистолен размер, ФИ – фракция на изтласкване, ФС – фракция на съкращение

Най-значими корелации се наблюдават между PAPm и изминатото разстояние на 6MWT ( $r = -0.58$ ), между PAPm и кислородната сатурация след 6MWT ( $r = -0.45$ ) и между ВАЦ и кислородната сатурация след 6MWT ( $r = 0.31$ ). PAPm корелира и със скалата на задуха (0.63), BODE индекса ( $r = 0.61$ ) и САТ ( $r = 0.60$ ), а също и с ФЕО1 ( $r = -0.43$ ).

### Ехокардиографски параметри и композитна крайна точка

64 пациенти (67%) от цялата група достигат композитната крайна точка (екзацербация или смърт).

Значимо по-голям процент от пациентите с отклонение в DT ( $\geq 250$  msec или  $< 150$  msec) достигат композитната крайна точка (86.2 срещу 58.2%,  $p = 0.0055$ ) (Фигура 2).



**Фигура 2. Kaplan-Meier крива за достигане на композитната крайна точка (време до екзацербация или смърт), стратифицирана по децелерационно време**

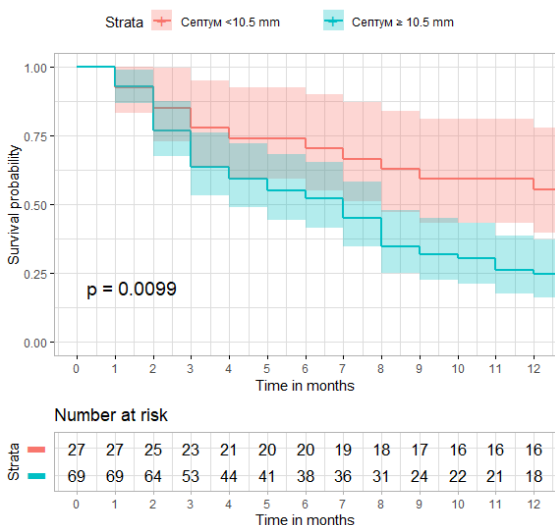
Съкращения: DT – децелерационно време

Пациентите с по-ниска ФС (78.6 срещу 57.4% при гранична точка 32%,  $p=0.049$ ) и уголемено ляво предсърдие (58.3 срещу

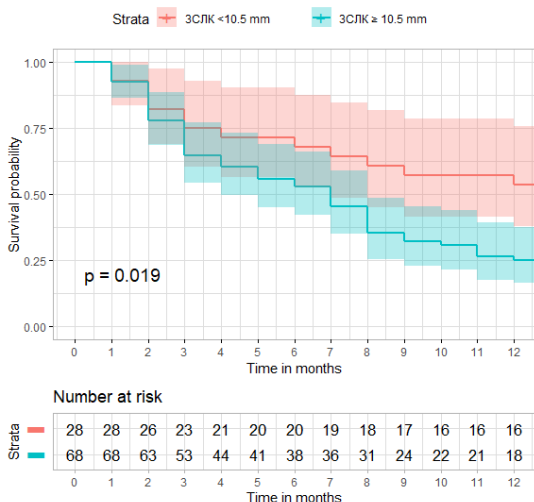
80.6% при гранична точка 36 mm,  $p=0.019$ ) достигат по-често композитната крайна точка.

Пациентите в двете групи не се различават по ТДР (оптимална гранична точка 43.5 mm,  $p = 0.09$ ), ТСР (оптимална гранична точка 29 mm,  $p = 0.07$ ), площ на ляво предсърдие, отношението Е/А, LAR и ФИ (всички  $p > 0.2$ ).

Пациентите с повишена дебелина на септума достигат значимо по-често композитната крайна точка (55.6 срещу 24.6% при гранична точка 10.5 mm,  $p=0.0099$ ) (Фигура 3). Подобен резултат се наблюдава и за ЗСЛК (55.6 срещу 24.6% при гранична точка 10.5 mm,  $p = 0.019$ ) (Фигура 4).



**Фигура 3. Kaplan-Meier крива за достигане на композитната крайна точка (време до екзацербация или смърт), стратифицирана по дебелина на септума**



**Фигура 4. Kaplan-Meier крива за достигане на композитната крайна точка (време до екзацербация или смърт), стратифицирана по дебелина на ЗСЛК**

Съкращения: ЗСЛК – задна стена на лява камера

#### Ехокардиографски параметри на лява камера и ляво предсърдие и честота на екзацербации

Пациентите с нормално DT (150-250 msec) имат значимо по-ниска честота на екзацербации (1.1 срещу 1.8,  $p = 0.01$ ). Това се дължи на значимо по-ниска честота на тежките екзацербации (0.29 срещу 0.85,  $p = 0.016$ ), като честотата на умерено-тежките екзацербации не се различава значимо (0.8 срещу 1.0,  $p = 0.32$ ).

Размерът на лявото предсърдие и LAR не са значими предиктори за честотата на екзацербациите като цяло или за умерено-тежките и тежките екзацербации. Пациентите с ФС  $\geq 32\%$  имат значимо по-малък брой екзацербации (1.04 срещу 1.61,  $p=0.025$ ).

Размерът на септума не е предиктор за общия брой екзацербации или за тежките екзацербации, но е предиктор за умерено-тежките екзацербации (0.52 срещу 1.0 при гранична точка 10.5 mm,  $p=0.01$ ). Аналогични са и резултатите за ЗСЛК (умерено-тежки екзацербации: 0.57 срещу 0.98 при гранична точка 10.5 mm,  $p=0.036$ ).

### Ехокардиографски параметри на лява камера и ляво предсърдие и качество на живот

Размер на ляво предсърдие  $<36$  mm или нормално DT е свързано с по-добро качество на живот, демонстрирано от по-нисък резултат от САТ (съответно 13.9 срещу 18.5,  $p=0.019$  и 14.3 срещу 18.6,  $p=0.048$ ). Аналогично, пациенти със септум  $\geq 10.5$  mm имат влошено качество на живот (16.8 срещу 12.6 точки на САТ,  $p=0.04$ ). Налице е тенденция за влошено качество на живот при пациенти с ФС  $<32\%$  (14 срещу 17.8 точки на САТ,  $p=0.0503$ ).

### Ехокардиографски параметри на лява камера и ляво предсърдие и белодробна функция

Въпреки че ФЕО1 е числено по-висок в групата с нормално DT, резултатите не достигат статистическа значимост (57.8 срещу 51.0%,  $p=0.099$ ). Няма значима разлика в белодробната функция при пациенти с повишен размер на ляво предсърдие и в зависимост от ФС, дебелината на септума, ЗСЛК или LAR (всички  $p > 0.05$ ).

### Шестминутен тест с ходене

След физическо натоварване (6MWT) най-значими корелации се наблюдават между PAPm и изминатото разстояние ( $r = -0.53$ ), между PAPm и кислородната сатурация след 6MWT ( $r = -0.48$ ) и между ВАЦ и кислородната сатурация след 6MWT ( $r = 0.46$ ). PAPm след 6MWT корелира и със скалата на задуха (0.62), BODE индекса ( $r = 0.6$ ) и САТ ( $r = 0.58$ ), а също и с ФЕО1 ( $r = -0.4$ ).

### Ехокардиографски параметри на дясна камера и дясно предсърдие *Честота на ПХ и резултати при едногодишно проследяване*

Разпространението на ПХ (PAPm в покой  $\geq 20$  mmHg) в цялата извадка е 52.1% (50/96). Пациентите с ПХ имат значимо пониски стойности на ФЕО1 (48.4 спрямо 63.7%,  $p < 0.0001$ ) и ФВК (70.6 спрямо 85.9%,  $p = 0.0006$ ). PAPm корелира значимо с ФЕО1 ( $r = -0.45$ ,  $p < 0.0001$ ) и ФВК ( $r = -0.34$ ,  $p = 0.0006$ ). PAPm в покой и след физическо натоварване също корелира с 6MWD (съответно  $r = -0.58$ ,  $p < 0.0001$ , и  $r = -0.52$ ,  $p < 0.0001$ ).

Шестдесет и четирима пациенти (67%) достигат композитната крайна точка, а 43 пациенти (45%) са чести екзацербатори (GOLD E) (Таблица 13). Пациентите с композитната крайна точка имат значимо по-високо PAPm в покой (24.3 спрямо 18.6 mmHg,  $p = 0.0001$ ) и след физическо натоварване (28.2 спрямо

22.2 mmHg,  $p = 0.001$ ). Честите екзацербации имат тенденция към по-високо PAPm в покой (21.1 спрямо 23.9 mmHg,  $p = 0.051$ ) и след физическо натоварване (27.8 спрямо 24.9 mmHg,  $p = 0.094$ ).

**Таблица 13. Резултати в зависимост от наличието на ПХ**

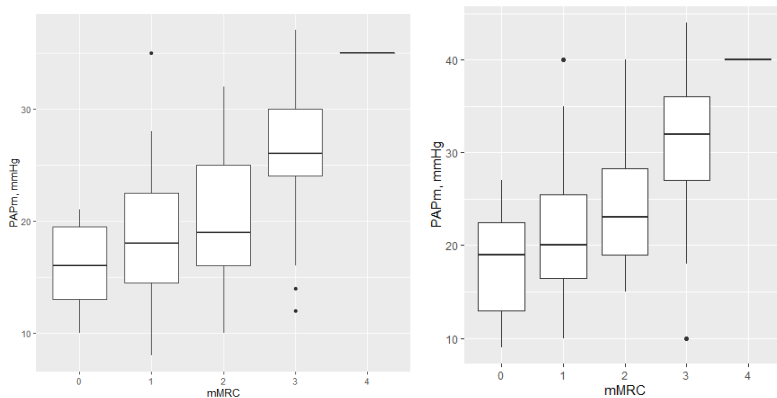
Крайна точка	Всички пациенти (n=96)	Пациенти без ПХ (n=46)	Пациенти с ПХ (n=50)	P-стойност за наличие срещу липса на ПХ
Умерено-тежки екзацербации след 1 година - средна стойност ( $\pm$ SD) - медиана (IQR)	0.86 $\pm$ 0.90 1 (0-1)	0.52 $\pm$ 0.76 0 (0-1)	1.16 $\pm$ 0.91 1 (1-1.75)	0.0004
Тежки екзацербации след 1 година - средна стойност ( $\pm$ SD) - медиана (IQR)	0.45 $\pm$ 0.87 0 (0-1)	0.36 $\pm$ 0.89 0 (0-0)	0.53 $\pm$ 0.84 0 (0-1)	0.358
Общ брой екзацербации	1.29 $\pm$ 1.21	0.89 $\pm$ 1.24	1.65 $\pm$ 1.07	0.002
Чести екзацербации	43 (44.8%)	14 (30.4%)	29 (58%)	0.006
Смъртни случаи	3 (3.1%)	2 (4.3%)	1 (2.0%)	0.94
Композитна крайна точка	64 (66.7%)	22 (47.8%)	42 (84%)	0.0004

Пациентите с ПХ в покой имат значимо по-голям брой екзацербации (1.65 спрямо 0.89,  $p = 0.002$ ). Разликите обаче са значими само за умерено-тежките екзацербации (1.16 спрямо 0.52,  $p = 0.0003$ ), но не и за тежките обостряния (0.53 спрямо 0.36,  $p = 0.36$ ). PAPm е значим предиктор само за умерено-тежките екзацербации.

Пациентите с PAPm  $\geq 30$  mmHg след физическо натоварване имат и значимо по-голям брой екзацербации (1.64 срещу 1.1,  $p = 0.026$ ), като по същия начин тези разлики са значими само за умерено-тежките екзацербации (1.21 срещу 0.67,  $p = 0.005$ ), но не и за тежките екзацербации (0.48 срещу 0.43,  $p = 0.76$ ).

## Пулмонална хипертония и качество на живот

Пациентите с ПХ са със значимо по-лошо качество на живот (24.6 срещу 18.1 точки на САТ теста,  $p < 0.0002$ ). Степента на диспнея корелира с PAPm ( $r = 0.54$ ,  $p < 0.0001$ ) (Фигура 5). Наличието на повече симптоми (по-висок брой точки на САТ) също корелира с PAPm ( $r = 0.61$ ,  $p < 0.0001$ ).



**Фигура 5. Средно налягане в пулмоналната артерия (PAPm) в покой (вляво) и след 6MWT (вдясно) в зависимост от mMRC групата**

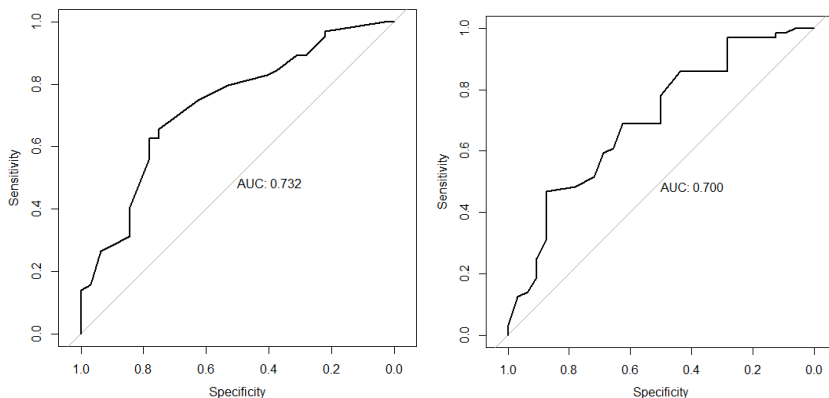
*Фактори, свързани с екзацербациите: еднофакторен и многофакторен анализ*

При еднофакторния анализ възрастта, пакетогодините, ФЕО1, ФВК, 6MWD, САТ резултатът, предходните екзацербации, PAPm и PAPm след усилие са значими предиктори за екзацербациите и/или композитната крайна точка. При многофакторния анализ обаче само предходните екзацербации остават значими във всички модели. ФВК е предиктор за умерено-тежки екзацербации и композитната крайна точка, а пакетогодините - за умерено-тежките екзацербации.

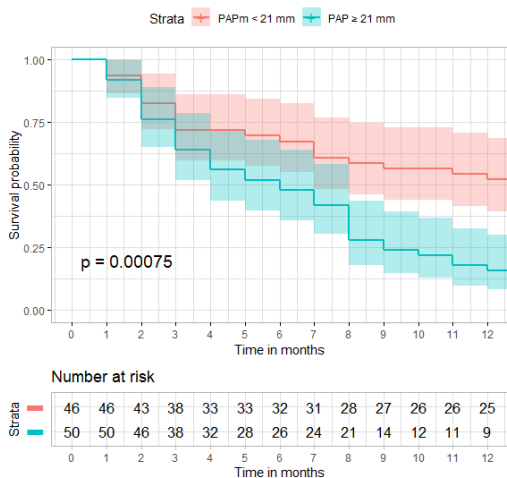
*Оптимални гранични стойности за PAPm в покой и след физическо натоварване*

Оптималните гранични точки, които осигуряват максимална стойност на Youden's J за композитната крайна точка, са 21 mmHg за PAPm в покой (AUC 0.732) и 30 mmHg след физическо натоварване (AUC 0.700). ROC кривите са показани на Фигура 6.

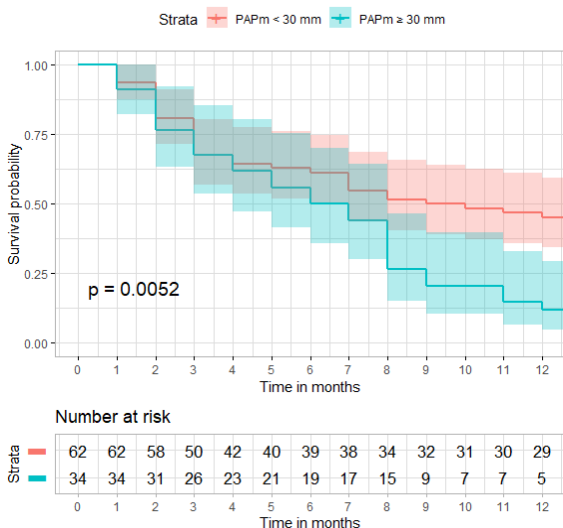
Пациентите с PAPm  $\geq 21$  mmHg са с повишен риск за развитие на композитната крайна точка (84% спрямо 47.8%,  $p < 0,0001$ , Фигура 7). Пациентите с PAPm  $\geq 30$  mmHg след 6MWT също са с повишен риск за развитие на композитната крайна точка (88.2% спрямо 54.8%,  $p=0,0002$ ) със специфичност 87.5% (Фигура 8).



**Фигура 6. ROC криви за средното налягане в пулмоналната артерия (PAPm) в покой (вляво) и след 6MWT (вдясно)**



**Фигура 7. Крива на преживяемост за достигане на композитната крайна точка (време до обостряне или смърт), стратифицирана според PAPm в покой ( $\geq 21$  спрямо  $< 21$  mmHg)**



**Фигура 8. Крива на преживяемост за достигане на композитната крайна точка (време до обостряне или смърт), стратифицирана според PAPm след 6MWT ( $\geq 30$  спрямо  $< 30$  mmHg)**

По отношение на честотата на екзацербации, оптималните гранични точки, които осигуряват максимална стойност на Youden's J, са същите - 21 mmHg за PAPm в покой и 30 mmHg след натоварване. Пациентите с PAPm  $\geq 21$  mmHg са с повишен риск за чести екзацербации (58.0% спрямо 30.4% пациенти,  $p = 0.012$ ). Въпреки това PAPm  $> 30$  mmHg след 6MWT не е значим предиктор за чести екзацербации (54.8% спрямо 40.0%,  $p = 0.25$ )

### Ехографски показатели

На Таблица 14 са представени използваните променливи, техният вид, начинът на кодиране и мерните единици.

**Таблица 14. Характеристики на променливите (ехографски показатели)**

Променлива	Вид	Начин на кодиране	Мерни единици	Средна стойност
Дясна диафрагма в покой	Числова	Цяло число	mm	48.1
Лява диафрагма в	Числова	Цяло число	mm	47.6

покой				
Дясна диафрагма след 6MWT	Числова	Цяло число	mm	41.8
Лява диафрагма след 6MWT	Числова	Цяло число	mm	41.1

В тази група няма липсващи стойности и допълнителна обработка на данните не се налага.

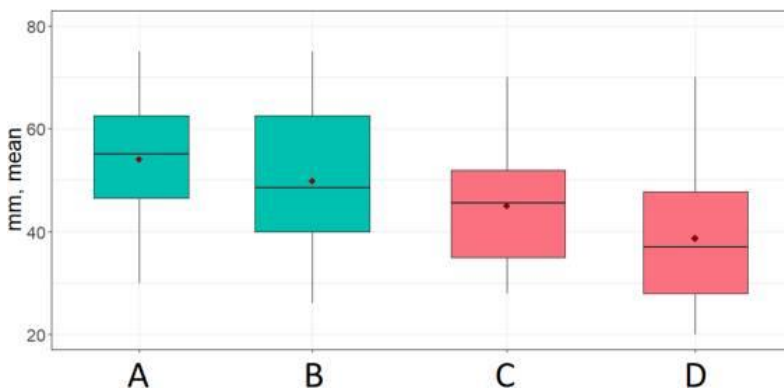
Средната диафрагмална подвижност в покой за дясната диафрагма е 48.1 mm (SD: 11.4 mm), а след 6MWT – 41.8 mm (SD: 14.6 mm) (Таблица 15). Средната диафрагмална подвижност за лявата диафрагма в покой е 47.6 mm (SD: 11.9 mm), а след 6MWT е 41.1 mm (SD: 15.5 mm). Резултатите за максималната подвижност между лява и дясна диафрагма не се различават значимо. С цел избягване на излишни повторения, ще бъдат представени само резултатите за подвижност на дясната диафрагма.

**Таблица 15. Основни ехографски параметри**

Характеристика	Средна стойност ± SD
Амплитуда на дясна диафрагма преди натоварване, mm	48.1 ± 11.4
Амплитуда на дясна диафрагма след натоварване, mm	41.8 ± 14.6
Амплитуда на лява диафрагма преди натоварване, mm	47.6 ± 11.9
Амплитуда на лява диафрагма след натоварване, mm	41.1 ± 15.5

#### Ехографски параметри и композитна крайна точка

64 пациенти (67%) от цялата група достигат композитната крайна точка (време до екзацербация или смърт). Средната диафрагмална подвижност в покой при пациентите без композитната крайна точка е по-висока (54.7 срещу 44.8,  $p < 0.0001$ ). Средната диафрагмална подвижност след 6MWT при пациентите без композитната крайна точка също е по-висока (50.2 срещу 37.7,  $p < 0.0001$ ). Съответните разлики от 9.9 mm (при покой) и 12.5 mm (след 6MWT) са значими (Фигура 9).



**Фигура 9. Максимална подвижност на дясната диафрагма в покой и след 6MWT.**

A - Максимална подвижност на дясната диафрагма в покой при пациенти без композитната крайна точка.

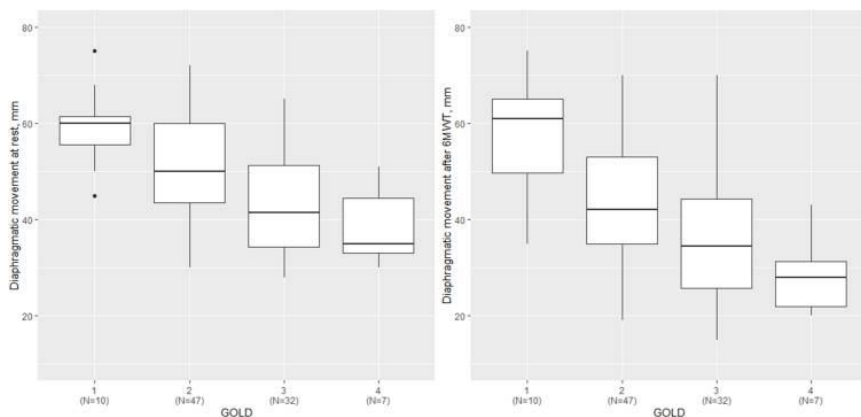
B - Максимална подвижност на дясната диафрагма след 6MWT при пациенти без композитната крайна точка.

C - Максимална подвижност на дясната диафрагма в покой при пациенти, достигащи композитната крайна точка.

D) - Максимална подвижност на дясната диафрагма след 6MWT при пациенти, достигащи композитната крайна точка.

$p < 0,0001$  за (A) спрямо (C) и (B) спрямо (D). Оцветеният правоъгълник представлява интерквartilния обхват (IQR). Хоризонталната черна линия в правоъгълника показва медианата. Червената точка показва средната стойност. Вертикалната черна линия обхваща минималните ( $Q1 - 1,5 \cdot IQR$ ) и максималните ( $Q3 + 1,5 \cdot IQR$ ) стойности (whiskers).

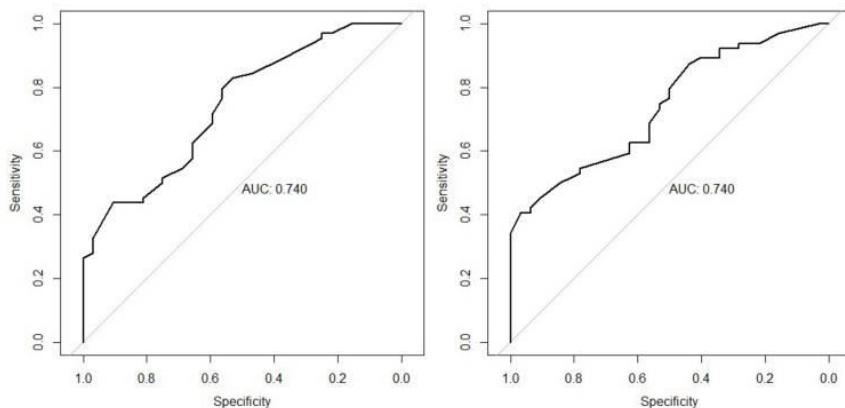
Диафрагмалната подвижност в покой корелира положително с ФЕО1 – 59.4 mm при GOLD I, 50.8 mm при GOLD II, 42.7 mm при GOLD III и 38.7 mm при GOLD IV (Фигура 10). Аналогични резултати се наблюдават и за диафрагмалната подвижност след 6MWT - 58.3 mm при GOLD I, 44.9 mm при GOLD II, 35.4 mm при GOLD III и 28.2 mm при GOLD IV (Фигура 10).



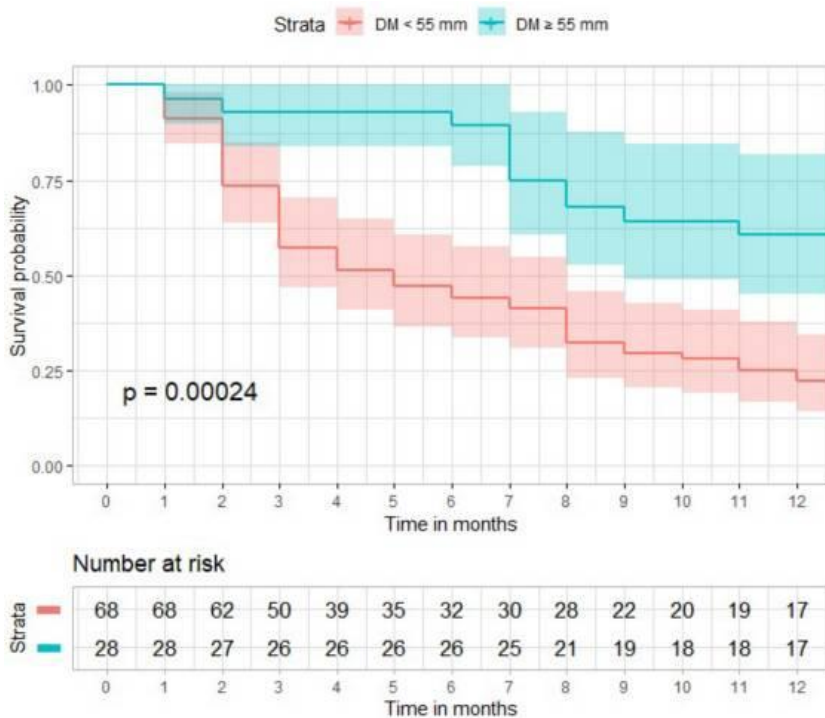
**Фигура 10. Диафрагмална подвижност в покой и след 6MWT в зависимост от стадия по GOLD.**

Белият правоъгълник представлява интерквартилния обхват (IQR). Хоризонталната черна линия в правоъгълника показва медианата. Вертикалната черна линия обхваща минималните ( $Q1 - 1,5 \cdot IQR$ ) и максималните ( $Q3 + 1,5 \cdot IQR$ ) стойности (whiskers).

Пациентите с достигната композитна крайна точка имат значимо по-ниска диафрагмална подвижност в покой. Оптималните гранични точки, които осигуряват максимална информативност, са 55 mm за диафрагмална подвижност в покой (точност 72,9%) и 53 mm след натоварване. ROC кривите са показани на Фигура 11. Пациентите с диафрагмална подвижност  $< 55$  mm са с повишен риск за достигане на композитната крайна точка (77,9% спрямо 39,3% от пациентите, HR: 1,98,  $p = 0,00024$ , Фигура 12).



**Фигура 11. ROC криви за диафрагмална подвижност в покой (вляво) и след 6MWT (вдясно).**

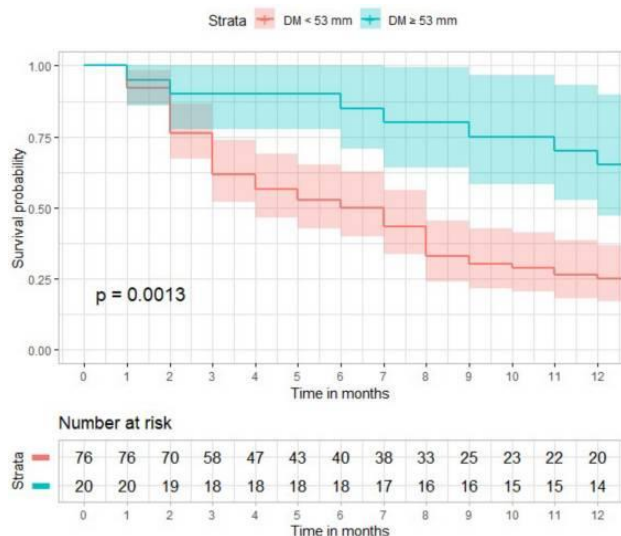


**Фигура 12. Kaplan-Meier крива за достигане на композитната крайна точка (време до екзацербация или смърт),**

**стратифицирана според диафрагмалната подвижност в покой ( $\geq 55$  срещу  $< 55$  mm).**

Съкращения: DM - диафрагмална подвижност.

Диафрагмалната подвижност след 6MWT притежава същата точност (72,9%). Пациентите с диафрагмална подвижност след 6MWT  $< 53$  mm също са с повишен риск за достигане на композитната крайна точка (75% срещу 35% от пациентите, HR: 2,14,  $p=0,0013$ ) (Фигура 13).

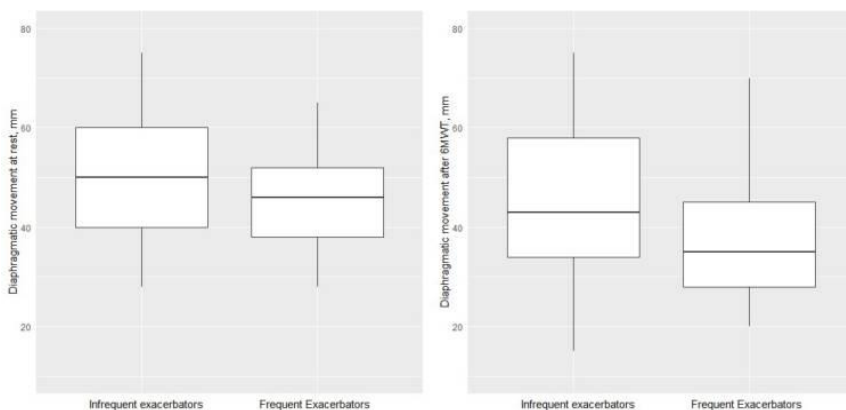


**Фигура 13. Kaplan-Meier крива за достигане на композитната крайна точка (време до екзацербация или смърт), стратифицирана според диафрагмалната подвижност след 6MWT ( $\geq 53$  срещу  $< 53$  mm).**

Съкращения: DM - диафрагмална подвижност.

Диафрагмалната подвижност в покой и след 6MWT е по-висока при пациентите без чести екзацербации (съответно  $50.0 \pm 12.2$  и  $44.9 \pm 15.6$  mm) в сравнение с пациентите с чести екзацербации по GOLD (съответно  $45.4 \pm 10.1$  и  $37.7 \pm 12.5$  mm,  $p = 0.047$  в покой и  $p = 0.01$  след 6MWT) (Фигура 14). Освен това пациентите с намалена диафрагмална подвижност имат по-висока честота на умерено-тежки екзацербации както в покой ( $1.06 \pm 0.93$  спрямо  $0.39 \pm 0.63$ ,  $p = 0.0001$ ), така и след 6MWT ( $0.97 \pm 0.89$

спрямо  $0.45 \pm 0.83$ ,  $p = 0.019$ ). В допълнение, пациентите с намалена диафрагмална подвижност след 6MWT имат повече тежки екзацербации ( $0.54 \pm 0.94$  спрямо  $0.1 \pm 0.31$ ,  $p = 0.0009$ ).

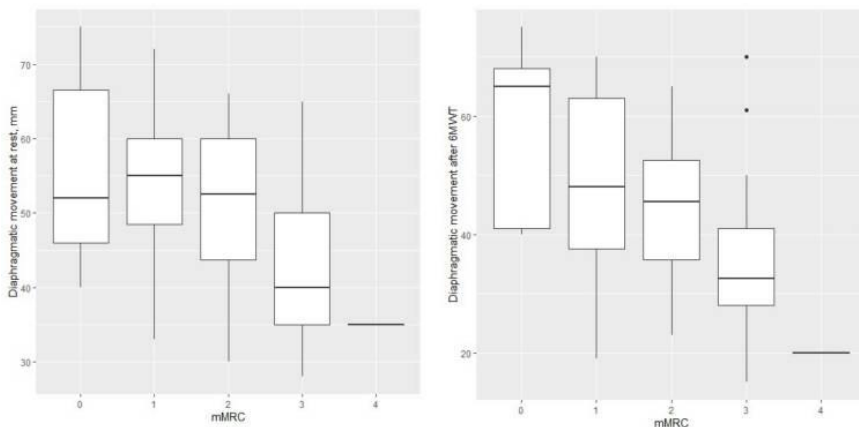


**Фигура 14. Диафрагмална подвижност в покой (вляво) и след 6MWT (вдясно) в зависимост от екзацербациите.**

За „чест екзацербатор“ се приема наличие на тежка екзацербация или  $\geq 2$  умерено-тежки екзацербации при едногодишно проследяване.

Белият правоъгълник представлява интерквартилния обхват (IQR). Хоризонталната черна линия в правоъгълника показва медианата. Вертикалната черна линия обхваща минималните ( $Q1 - 1,5 \cdot IQR$ ) и максималните ( $Q3 + 1,5 \cdot IQR$ ) стойности (whiskers).

Степента на диспнея е свързана с диафрагмалната подвижност в покой (56.4, 53.5, 50.8, 42.3 и 35.0 mm съответно за резултати по mMRC „0“, „1“, „2“, „3“ и „4“) и след 6MWT (56.9, 47.9, 44.2, 34.5 и 20.0 mm съответно за резултати по mMRC „0“, „1“, „2“, „3“ и „4“) (Фигура 15). Наличието на повече симптоми (CAT  $\geq 10$ ) също е свързано с намалена диафрагмална подвижност след усилие (38.1 срещу 49.0 mm,  $p = 0.0009$ , 87,3% специфичност).



**Фигура 15. Диафрагмална подвижност в покой (вляво) и след 6MWT (вдясно) според резултата от mMRC.**

Белият правоъгълник представлява интерквартилния обхват (IQR). Хоризонталната черна линия в правоъгълника показва медианата. Вертикалната черна линия обхваща минималните ( $Q1 - 1,5 \cdot IQR$ ) и максималните ( $Q3 + 1,5 \cdot IQR$ ) стойности (whiskers).

### Електрокардиографски показатели

На Таблица 16 са представени използваните променливи, техният вид, начинът на кодиране и мерните единици.

**Таблица 16. Характеристики на променливите (електрокардиографски показатели)**

Променлива	Вид	Начин на кодиране	Мерни единици	Средна стойност
Сърдечна честота	Числова	Цяло число	$\text{min}^{-1}$	80.7
Височина на рП-вълната	Закръгляване до първия знак след десетичната запетая	Цяло число	mm	1.4
Други отклонения	Текстова	Структуриран текст	-	-

В тази група няма липсващи стойности и допълнителна обработка на данните не се налага.

Средната сърдечна честота в покой е 80.7/min ( $\pm 13.1$ /min) (**Error! Reference source not found.**). Честотата на пациентите с тахикардия ( $>90$ /min) е 25% (24/96), а на тези с брадикардия ( $<60$ /min) е 3.1% (3/96). Предсърдно мъждене имат 7.3% (7/96), а десен бедрен блок – 8.3% (8/96). Белези на левокамерно обременяване имат 31.3% (30/96), а на деснокамерно обременяване/cog pulmonale – 15.6% (15/96). Средната височина на р-вълната, измерена във II отвеждане, е 1.5 mm ( $\pm 0.5$  mm). Данни за преживян инфаркт имат 4 пациента (4.2%).

Сърдечната честота не показва зависимост с тежестта на ХОББ – средни стойности съответно 82.9/min (ФЕО1  $>80\%$ ), 86.8/min (ФЕО1 между 50 и 80%), 78.2/min (ФЕО1 между 30 и 50%) и 81.9/min (ФЕО1  $<30\%$ ). Също така сърдечната честота не е самостоятелен прогностичен фактор за риск от екзацербации или смърт при едногодишно проследяване (всички  $p > 0.1$ ).

Най-висока корелация сърдечната честота в покой показва с десатурацията след 6MWT ( $r = 0.34$ ) и кислородната сатурация след 6MWT ( $r = -0.24$ ). Височината на р-вълната при пациентите без предсърдно мъждене показва най-висока корелация със сърдечната честота в покой ( $r = 0.49$ ), ФЕО1/ФВК отношението ( $r = -0.39$ ) и давността на ХОББ ( $r = 0.36$ ). Трябва да се отбележи, че подобно на сърдечната честота в покой, тук също се наблюдава корелация с кислородната сатурация след 6MWT ( $r = -0.35$ ) и кислородната десатурация след 6MWT ( $r = 0.33$ ).

### Оценка на физическия капацитет

На Таблица 17 са представени използваните променливи, техният вид, начинът на кодиране и мерните единици.

**Таблица 17. Характеристики на променливите (оценка на физическия капацитет)**

Променлива	Вид	Начин на кодиране	Мерни единици	Средна стойност
Продължителност на 6MWT	Числова	Цяло число	min	5.95
Изминато разстояние на 6MWT	Числова	Цяло число	m	372.3
Кислородна сатурация преди 6MWT	Числова	Цяло число	%	94.6
Кислородна сатурация след 6MWT	Числова	Цяло число	%	90.1

Десатурационен индекс /разликата между кислородната сатурация в края спрямо началото на теста/	Числова	Цяло число	%	4.5
BODE индекс	Числова	Цяло число	Точки	3.1

В тази група няма липсващи стойности и допълнителна обработка на данните не се налага.

Почти всички пациенти (96.9%, 93/96) са завършили успешно 6MWT. Средното изминато разстояние е  $372.3 \pm 110.2$  m. Средната кислородна сатурация в покой е  $94.6 \pm 2.7\%$ , а след 6MWT –  $90.1\% \pm 5.9\%$ , което прави десатурационен индекс от  $4.5 \pm 4.5\%$ . 5 пациенти (5.2%) са имали дихателна недостатъчност в покой (кислородна сатурация  $<90\%$ ), докато такава се наблюдава при 37.5% от пациентите след 6MWT (36/96).

Най-висока корелация изминатото разстояние на 6MWT показва с BODE индекса ( $r = -0.79$ ), скалата на задуха ( $r = -0.71$ ) и SAT ( $r = -0.7$ ). Интересна корелация се наблюдава и с PAPm в покой ( $r = -0.58$ ) и PAPm след 6MWT ( $r = -0.53$ ). Изминатото разстояние корелира още с възрастта ( $r = -0.37$ ), пакетогодините тютюнопушене ( $r = -0.33$ ), индекса на Charlson ( $r = -0.36$ ), ФЕО1 ( $r = 0.48$ ) и подвижността на диафрагмата в покой ( $r = 0.46$ ) и след 6MWT ( $r = 0.54$ ).

### Едногодишно проследяване

На Таблица 18 са представени използваните променливи, техният вид, начинът на кодиране и мерните единици. На Таблица 19 са представени липсващите стойности и последващото действие за тяхната обработка.

**Таблица 18. Характеристики на променливите (едногодишно проследяване)**

Променлива	Вид	Начин на кодиране	Мерни единици	Средна стойност
Умерено-тежки екзацербации след 6 месеца	Числова	Цяло число	брой	0.41
Тежки екзацербации след 6 месеца	Числова	Цяло число	брой	0.16

Умерено-тежки екзацербации след 12 месеца	Числова	Цяло число	брой	0.86
Тежки екзацербации след 12 месеца	Числова	Цяло число	брой	0.45

**Таблица 19. Липсващи стойности (едногодишно проследяване)**

Променлива	Липсващи стойности	Действие
Умерено-тежки екзацербации след 6 месеца	2	Без заместване (липсващи стойности поради летален изход)
Тежки екзацербации след 6 месеца	2	Без заместване (липсващи стойности поради летален изход)
Умерено-тежки екзацербации след 12 месеца	3	Без заместване (липсващи стойности поради летален изход)
Тежки екзацербации след 12 месеца	3	Без заместване (липсващи стойности поради летален изход)

## 4.2. Машинно обучение за прогнозиране на екзацербации

### Наличие на екзацербация

66.7% (64/96) от пациентите са имали екзацербация в предходната година. С цел възпроизводимост на резултатите е използван използван *seed* за случайните числа (*set.seed(100)*).

### Нулев модел

За нулев модел (модел без предиктори) ще бъде използван отговорът с по-висока честота, т.е. моделът прогнозира, че всички пациенти с ХОББ ще имат екзацербация. Нулевият модел класифицира правилно 65.5% (19/29) от тестовите наблюдения (Таблица 20).

**Таблица 20. Точност на нулевия модел**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	0	10
Екзацербатор	0	19
NPV/PPV	0	0.655

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

### Логистична регресия

Проведена е двупосочна (direction = "both") логистична регресия за прогнозиране на наличието на поне една екзацербация при едногодишно проследяване. Резултатите от коефициентите са представени в Таблица 21.

**Таблица 21. Коефициенти на модела с логистична регресия (подредени по абсолютна стойност)**

Параметър	Стойност
Хематокрит	-11033
Intercept (пресечна точка)	-1012
Наличие на анамнеза за бронхиална астма	797
Наличие на анамнеза за хроничен бронхит	438
Симптом: свирене в гърдите	-206
TSP	48
Левкоцити	45
Възраст	40
Моноцити, %	35
ФЕО1 преди бронходилататор, %	-32
Хемоглобин	19
ФЕО1 след бронходилататор, %	16

Съкращения: TSP – телесистолен размер, ФЕО1 – форсиран едносекунден обем за първата секунда

Забележка: положителните стойности означават по-висок риск

Моделът класифицира правилно 65.5% (19/29) от тестовите наблюдения (Таблица 22).

**Таблица 22. Точност на модела с линейна регресия**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	3	7
Екзацербатор	3	16
NPV/PPV	0.5	0.7

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

## Метод на $K$ -най-близките съеди (KNN)

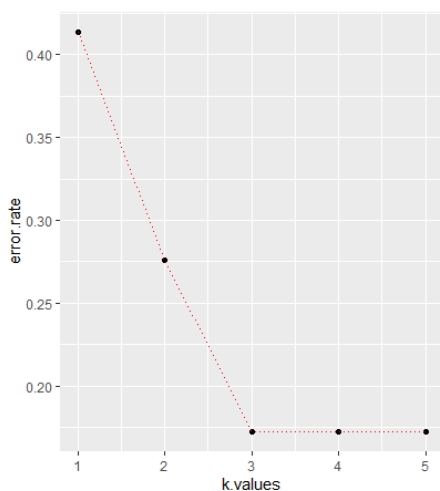
Създаден е модел с KNN алгоритъм за прогнозиране на наличието на поне една екзацербация при едногодишно проследяване.

При  $k = 1$  точността на модела е 58.6% (17/29) (Таблица 23). Моделът има най-висока точност при  $k = 3$  – 82.8% (24/29) (Фигура 16) (Таблица 24).

**Таблица 23. Точност на KNN модела при  $k = 1$**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	5	5
Екзацербатор	7	12
NPV/PPV	0.42	0.71

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност



**Фигура 16. Съпоставяне между  $K$ -стойностите и точността**

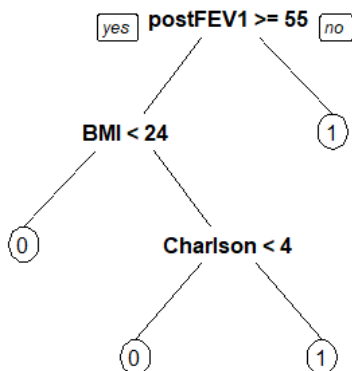
**Таблица 24. Точност на KNN модела при  $k = 3$**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	6	4
Екзацербатор	1	18
NPV/PPV	0.86	0.82

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

## Дърво на решението

Създаден е модел с дърво на решението за прогнозиране на наличието на поне една екзацербация при едногодишно проследяване (Фигура 17).



**Фигура 17. Модел на дърво на решението за прогнозиране на екзацербация**

Съкращения: Charlson = Charlson индекс на коморбидност; postFEV1 – ФЕО1 (%) след бронходилатация

Моделът класифицира правилно 79.3% (23/29) от тестовите наблюдения (Таблица 25).

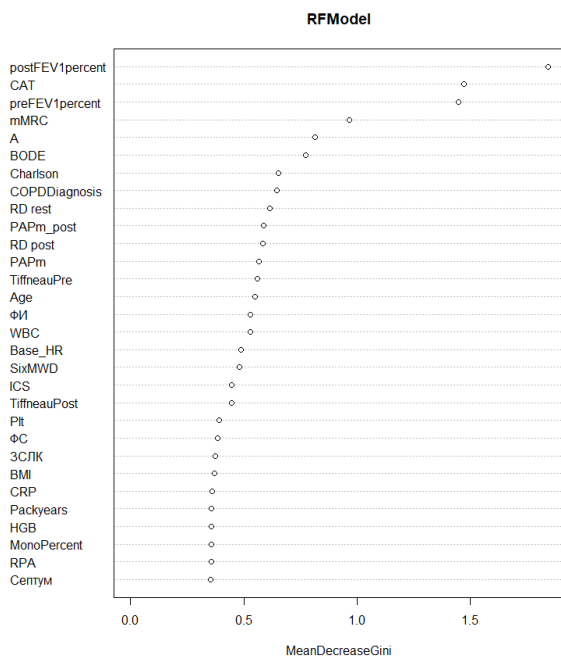
**Таблица 25. Точност на модела с дърво на решението**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	6	4
Екзацербатор	2	17
NPV/PPV	0.75	0.81

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

## Random Forest модел

Създаден е Random Forest модел за прогнозиране на наличието на поне една екзацербация при едногодишно проследяване. Тежестта на отделните предиктори е показана на Фигура 18.



**Фигура 18. Random Forest модел за прогнозиране на екзацербация**

Съкращения: postFEV1percent – ФЕО1 (%) след бронходилататор, preFEV1percent - ФЕО1 (%) преди бронходилататор, A - скорост на А вълната (ехокардиография), Charlson – Charlson индекс на коморбидност, COPDDiagnosis – давност на диагнозата ХОББ, RD rest – подвижност на дясна диафрагма в покой, PAPm\_post – средно налягане в а. pulmonalis след 6MWT, RD post - подвижност на дясна диафрагма след 6MWT, PAPm - средно налягане в а. pulmonalis в покой, TiffneauPre – ФЕО1/ФВК преди бронходилатация, Age - възраст, ФИ – фракция на изтласкване (ехокардиография), WBC - левкоцити, Base\_HR – сърдечна честота в покой, SixMWD – изминато разстояние на 6MWT, ICS – употреба на инхалаторни кортикостероиди, TiffneauPost - ФЕО1/ФВК след бронходилатация, Plt - тромбоцити, ФС – фракция на съкращение (ехокардиография), ЗСЛК – задна стена на лява камера (ехокардиография), Packyears – пакетогодини тютюнопушене, HGB - хемоглобин, MonoPercent – моноцити (%), RPA - диаметър на дясна пулмонална артерия (ехокардиография), септум – дебелина на септума (ехокардиография)

Моделът класифицира правилно 82.8% (24/29) от тестовите наблюдения (Таблица 26).

**Таблица 26. Точност на Random Forest модела**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	5	5
Екзацербатор	0	19
NPV/PPV	1.0	0.79

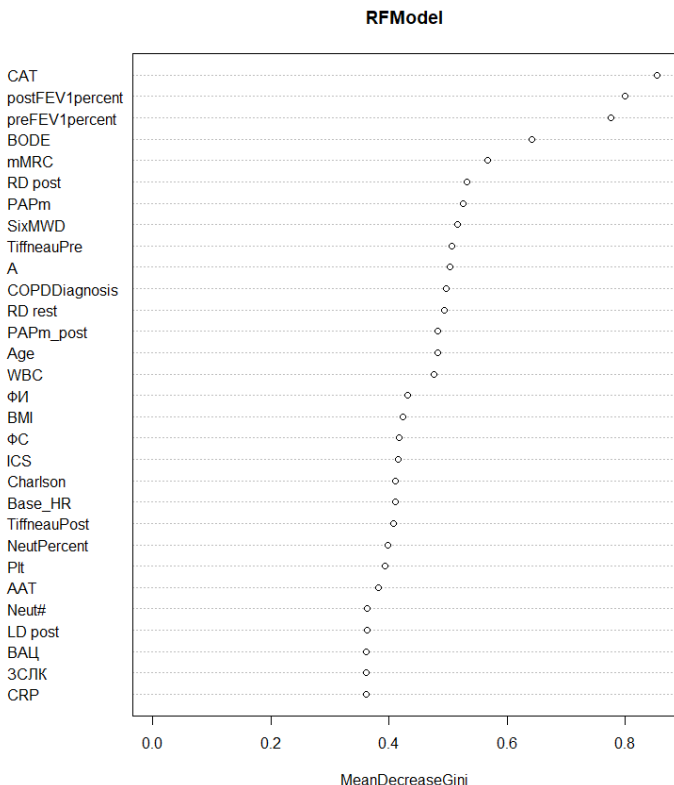
NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

Търсенето на оптималните хиперпараметри показва най-висока точност на модела при  $mtry = 2$  (брой на променливите, които се избират на случаен принцип като кандидати при всяко разделяне) и  $ntree = 2000$  (брой на дърветата). Моделът класифицира правилно 82.8% (24/29) от тестовите наблюдения (Таблица 27). Тежестта на отделните предиктори е показана на Фигура 19.

**Таблица 27. Точност на оптимизирания Random Forest модел**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	5	5
Екзацербатор	0	19
NPV/PPV	1.0	0.79

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност



**Фигура 19. Оптимизиран Random Forest модел за прогнозиране на екзацербация**

Съкращения: postFEV1percent – ФЕО1 (%) след бронходилататор, preFEV1percent - ФЕО1 (%) преди бронходилататор, RD post - подвижност на дясна диафрагма след 6MWT, PAPm - средно налягане в а. pulmonalis в покой (ехокардиография), SixMWD – изминато разстояние на 6MWT, TiffneauPre – ФЕО1/ФВК преди бронходилатация, А - скорост на А вълната (ехокардиография), COPDDiagnosis – давност на диагнозата ХОББ, RD rest – подвижност на дясна диафрагма в покой, PAPm\_post – средно налягане в а. pulmonalis след 6MWT, Age - възраст, WBC - левкоцити, ФИ – фракция на изтласкване (ехокардиография), ФС – фракция на съкращение (ехокардиография), ICS – употреба на инхалаторни кортикостероиди, Charlson – Charlson индекс на коморбидност, Base\_HR – сърдечна честота в покой, TiffneauPost - ФЕО1/ФВК след бронходилатация, NeutPercent – неутрофили (%), Plt - тромбоцити, ААТ – алфа1 антитрипсин, Neut# - неутрофили (абсолютен брой), LD post - подвижност на лява диафрагма след 6MWT, ВАЦ – време

на акцелерация (ехохардиография), ЗСЛК – задна стена на лява камера (ехохардиография).

### **Naive Bayes модел**

Създаден е Naive Bayes модел за прогнозиране на наличието на поне една екзацербация при едногодишно проследяване. Моделът класифицира правилно 82.8% (24/29) от тестовите наблюдения (Таблица 28).

**Таблица 28. Точност на Naive Bayes модела**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	5	5
Екзацербатор	0	19
NPV/PPV	1.0	0.79

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

### **Невронни мрежи (Neural Networks)**

Създаден е модел с невронна мрежа за прогнозиране на наличието на поне една екзацербация при едногодишно проследяване. Моделът класифицира правилно 65.5% (19/29) от тестовите наблюдения (Таблица 29).

**Таблица 29. Точност на модела с невронна мрежа**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	8	2
Екзацербатор	8	11
NPV/PPV	0.5	0.85

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

### **Поддържаща векторна машина (SVM)**

Създаден е SVM модел за прогнозиране на наличието на поне една екзацербация при едногодишно проследяване. Моделът класифицира правилно 79.3% (23/29) от тестовите наблюдения (Таблица 30).

**Таблица 30. Точност на SVM модела**

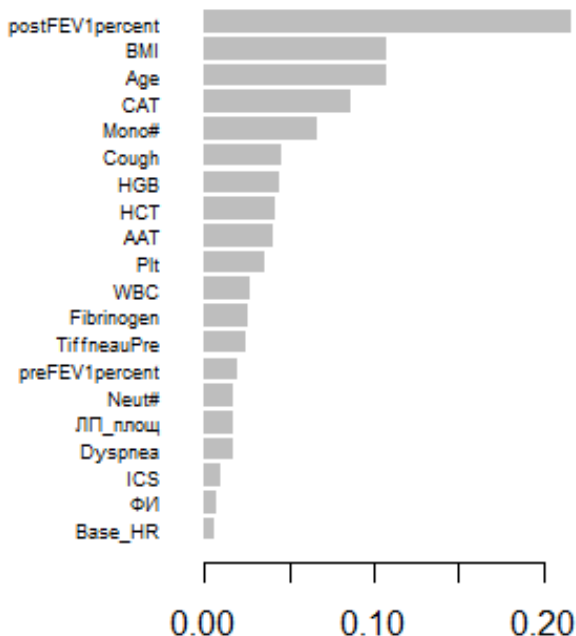
	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	5	5
Екзацербатор	1	18

NPV/PPV	0.83	0.78
---------	------	------

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

## XGBoost

Създаден е XGBoost модел за прогнозиране на наличието на поне една екзацербация при едногодишно проследяване. Тежестта на отделните предиктори е показана на Фигура 20.



### Фигура 20. Важност на характеристиките в XGBoost модела

Съкращения: postFEV1percent – ФЕО1 (%) след бронходилататор, Age - възраст, Mono# - моноцити (абсолютен брой), Cough – симптом: кашлица, HGB – хемоглобин, HCT – хематокрит, AAT – алфа1 антитрипсин, Plt - тромбоцити, WBC – левкоцити, TiffneauPre – ФЕО1/ФВК преди бронходилатация, preFEV1percent - ФЕО1 (%) преди бронходилататор, Neut# - неутрофили (абсолютен брой), ЛП\_плоч – площ на ляво предсърдие (ехокардиография), Dyspnea – симптом: задух, ICS – употреба на инхалаторни кортикостероиди, ФИ – фракция на изтласкване (ехокардиография), Base\_HR – сърдечна честота в покой

Моделът класифицира правилно 79.3% (23/29) от тестовите наблюдения (Таблица 31).

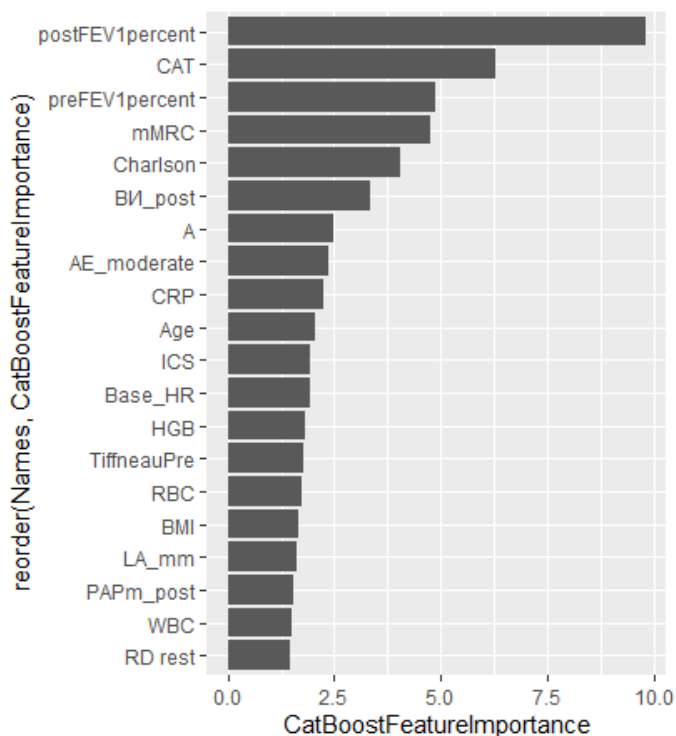
**Таблица 31. Точност на XGBoost модела**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	6	4
Екзацербатор	2	17
NPV/PPV	0.75	0.81

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

## CatBoost

Създаден е CatBoost модел за прогнозиране на наличието на поне една екзацербация при ежегодно проследяване. Тежестта на отделните предиктори е показана на Фигура 21.



**Фигура 21. Важност на характеристиките в CatBoost модела**

Съкращения: postFEV1percent – ФЕО1 (%) след бронходилататор, preFEV1percent - ФЕО1 (%) преди бронходилататор, Charlson – Charlson индекс на коморбидност, VI\_post – време на изтласкване след 6MWT

(ехохардиография), A - скорост на A вълната (ехохардиография), AE\_moderate – брой умерено-тежки екзацербации в предходната година, Age - възраст, ICS – употреба на инхалаторни кортикостероиди, Base\_HR – сърдечна честота в покой, HGB – хемоглобин, TiffneauPre – ФЕО1/ФВК преди бронходилатация, RBC – еритроцити, LA\_mm – размер на ляво предсърдие (ехохардиография), PAPm\_post – средно налягане в а. pulmonalis след 6MWT (ехохардиография), WBC - левкоцити, RD\_rest - подвижност на лява диафрагма в покой.

Моделът класифицира правилно 79.3% (23/29) от тестовите наблюдения (Таблица 32).

**Таблица 32. Точност на CatBoost модела**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	4	6
Екзацербатор	0	19
NPV/PPV	1.0	0.76

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

### **Ансамблов модел (Ensemble model)**

Ансамбловият модел за прогнозиране на наличието на поне една екзацербация при едногодишно проследяване включва трите самостоятелни модела с най-добри резултати. Това са KNN (82.8%), Random Forest (82.8%) и Naïve Bayes (82.8%). Гласуването е осъществено с просто мнозинство – избира се категорията с повече гласове (т.е.  $\geq 2$ ). Моделът класифицира правилно 82.8% (24/29) от тестовите наблюдения (Таблица 33).

**Таблица 33. Точност на ансамбловия модела**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	5	5
Екзацербатор	0	19
NPV/PPV	1.0	0.79

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

### **4.3. Риск по GOLD**

При едногодишно проследяване 44.8% (43/96) от пациентите в изследваната група попадат в група Е по GOLD (т.е.

са имали тежка екзацербация или  $\geq 2$  умерено-тежки екзацербации). С цел възпроизводимост на резултатите ще бъде използван *seed* за случайните числа (*set.seed(100)*)

### Нулев модел (GOLD модел)

Нулевият модел взема под внимание екзацербациите през предходната година. Съгласно препоръките на GOLD при пациентите, които в предходната година са били в група Е, е прогнозиран висок риск за екзацербации. Аналогично при пациентите, които са били в група А или В в предходната година, е прогнозиран нисък риск от екзацербации. Моделът класифицира правилно 58.6% (17/29) от тестовите наблюдения (Таблица 34).

**Таблица 34. Точност на нулевия модел (GOLD препоръки)**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	6	10
Екзацербатор	2	11
NPV/PPV	0.75	0.52

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

### Логистична регресия

Проведена е двупосочна (*direction = "both"*) логистична регресия за прогнозиране на пациенти с висок риск от екзацербации (група Е по GOLD). Резултатите от коефициентите са представени в Таблица 35.

**Таблица 35. Коефициенти на модела с логистична регресия (подредени по абсолютна стойност)**

Параметър	Стойност
Хематокрит	-8450
Intercept (пресечна точка)	2038
Анамнеза за хиперурикемия	503
Анамнеза за бронхиектазии	322
Анамнеза за предходна тежка екзацербация	274
Терапия с инхалаторни кортикостероиди	266
Мъжки пол	-245
mMRC резултат	131
ТСР на лява камера	-123
ТДР на лява камера	72
ФИ на лява камера	-46

Хемоглобин	33
ФЕО1/ФВК преди бронходилататор	-5

Забележка: положителните стойности означават повишен риск

Моделът класифицира правилно 75.9% (22/29) от тестовите наблюдения (Таблица 36).

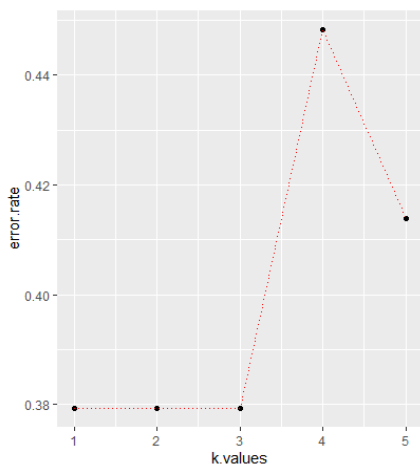
**Таблица 36. Точност на модела с логистична регресия**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	12	4
Екзацербатор	3	10
NPV/PPV	0.8	0.71

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

### Метод на *K*-най-близките съседи (KNN)

Създаден е модел с KNN алгоритъм за прогнозиране на пациенти с висок риск от екзацербации (група Е по GOLD). При  $k = 1$  точността на модела е 62.1% (18/29). Моделът има най-висока и еднаква точност при  $k = 1$ ,  $k = 2$  и  $k = 3$  (Фигура 22) (Таблица 37).



**Фигура 22. Съпоставяне между *K*-стойностите и точността**

**Таблица 37. Точност на KNN модела при  $k = 1$**

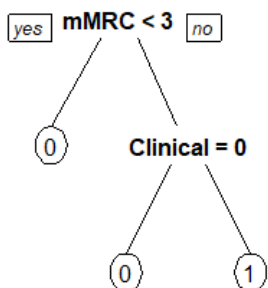
	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	10	6
Екзацербатор	5	8

NPV/PPV	0.67	0.57
---------	------	------

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

### Дърво на решението

Създаден е модел с дърво на решението за прогнозиране на пациенти с висок риск от екзацербации (група Е по GOLD) (Фигура 23).



**Фигура 23. Модел на дърво на решението за прогнозиране на пациенти с висок риск от екзацербации (група Е по GOLD)**

Забележка: Clinical = 0 означава клиничен фенотип на неекзацербатор в предходната година.

Моделът класифицира правилно 69.0% (20/29) от тестовите наблюдения (Таблица 38).

**Таблица 38. Точност на модела с дърво на решението**

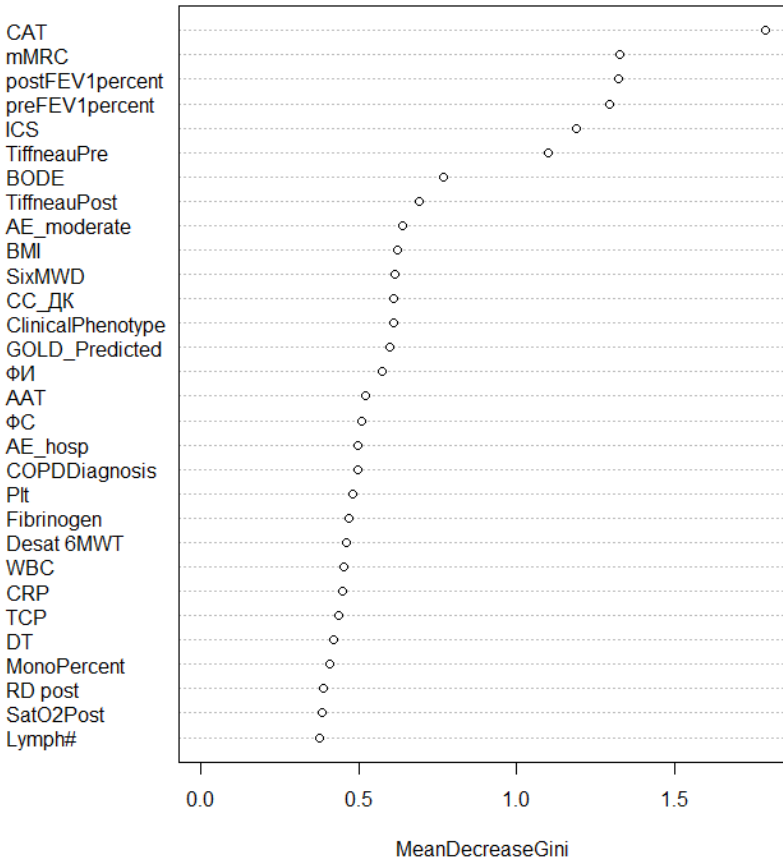
	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	13	3
Екзацербатор	6	7
NPV/PPV	0.68	0.7

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

### Random Forest модел

Създаден е Random Forest модел за прогнозиране на пациенти с висок риск от екзацербации (група Е по GOLD). Тежестта на отделните предиктори е показана на Фигура 24.

## RFModel



**Фигура 24. Random Forest модел за прогнозиране на пациенти с висок риск от екзацербации (група E по GOLD)**

Съкращения: postFEV1percent – ФЕО1 (%) след бронходилататор, preFEV1percent - ФЕО1 (%) преди бронходилататор, ICS – употреба на инхалаторни кортикостероиди, TiffneauPre – ФЕО1/ФВК преди бронходилатация, BODE – BODE индекс, TiffneauPost - ФЕО1/ФВК след бронходилатация, AE\_moderate – умерено-тежки екзацербации в предходната година, SixMWD – изминато разстояние на 6MWT, CC\_ДК – свободна стена на дясна камера (ехохардиография), ClinicalPhenotype – клиничен фенотип, GOLD\_Predicted – риск за екзацербация по GOLD (група E), ФИ – фракция на изтласкване (ехохардиография), AAT – алфа1 антитрипсин, фс – фракция на съкращение (ехохардиография), AE\_hosp - тежки екзацербации в предходната година, COPDDiagnosis – давност на диагнозата ХОББ, Plt - тромбоцити, Desat 6MWT – десатурационен индекс

след 6MWT, WBC - левкоцити, TSP – телесистолен размер на лява камера (ехокардиография), DT – децелерационно време (ехокардиография), MonoPercent – моноцити (%), RD post - подвижност на дясна диафрагма след 6MWT, SatO2Post – сатурация след 6MWT, Lymph# - лимфоцити (абсолютен брой)

Моделът класифицира правилно 62.1% (18/29) от тестовите наблюдения (Таблица 39).

**Таблица 39. Точност на Random Forest модела**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	11	5
Екзацербатор	6	7
NPV/PPV	0.65	0.58

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

Търсенето на оптималните параметри показва най-висока точност на модела при  $mtry = 9$  (брой на променливите, които се избират на случаен принцип като кандидати при всяко разделяне) и  $ntree = 1000$  (брой на дърветата). Моделът класифицира правилно 69.0% (20/29) от тестовите наблюдения (Таблица 40).

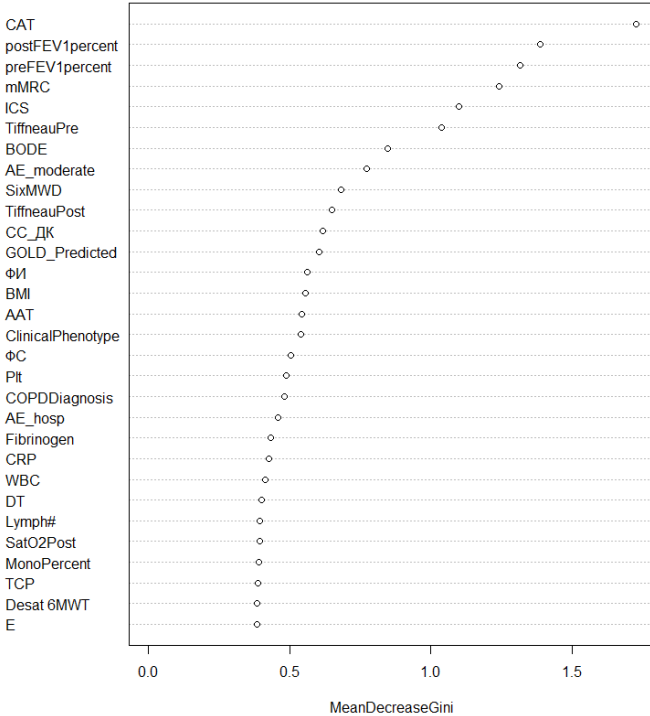
**Таблица 40. Точност на оптимизирания Random Forest модел**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	12	4
Екзацербатор	5	8
NPV/PPV	0.71	0.67

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

Тежестта на отделните предиктори е показана на Фигура 25.

RFModel



**Фигура 25. Оптимизиран Random Forest модел за прогнозиране на пациенти с висок риск от екзацербации (група E по GOLD)**

Съкращения: postFEV1percent – ФЕО1 (%) след бронходилататор, preFEV1percent - ФЕО1 (%) преди бронходилататор, ICS – употреба на инхалаторни кортикостероиди, TiffneauPre – ФЕО1/ФВК преди бронходилатация, BODE – BODE индекс, AE\_moderate – умерено-тежки екзацербации в предходната година, SixMWD – изминато разстояние на 6MWT, TiffneauPost - ФЕО1/ФВК след бронходилатация, CC\_ДК – свободна стена на дясна камера (ехокардиография), GOLD\_Predicted – риск за екзацербация по GOLD (група E), ФИ – фракция на изтласкване (ехокардиография), AAT – алфа1 антитрипсин, ClinicalPhenotype – клиничен фенотип, ФС – фракция на съкращение, Plt - тромбоцити, COPDDiagnosis – давност на диагнозата ХОББ, AE\_hosp - тежки екзацербации в предходната година, WBC - левкоцити, DT – децелерационно време (ехокардиография), Lymph# - лимфоцити (абсолютен брой), SatO2Post – сатурация след 6MWT, MonoPercent – моноцити (%), TCP – телесистолен размер на лява камера (ехокардиография), Desat 6MWT – десатурационен индекс след 6MWT, E – скорост на E вълната (ехокардиография).

## Naive Bayes модел

Създаден е Naive Bayes модел за прогнозиране на пациенти с висок риск от екзацербации (група Е по GOLD). Моделът класифицира правилно 55.2% (16/29) от тестовите наблюдения (Таблица 41).

**Таблица 41. Точност на Naive Bayes модела**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	14	2
Екзацербатор	11	2
NPV/PPV	0.56	0.5

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

## Невронни мрежи (Neural Networks)

Създаден е модел с невронна мрежа за прогнозиране на пациенти с висок риск от екзацербации (група Е по GOLD). Моделът класифицира правилно 62.1% (18/29) от тестовите наблюдения (Таблица 42).

**Таблица 42. Точност на модела с невронна мрежа**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	11	5
Екзацербатор	6	7
NPV/PPV	0.65	0.58

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

## Поддържаща векторна машина (SVM)

Създаден е SVM модел за прогнозиране на пациенти с висок риск от екзацербации (група Е по GOLD). Моделът класифицира правилно 55.2% (16/29) от тестовите наблюдения (Таблица 43).

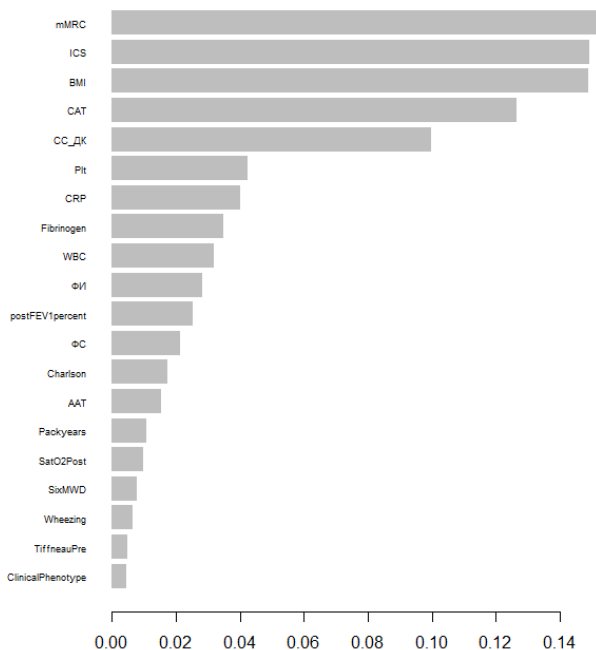
**Таблица 43. Точност на SVM модела**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	9	7
Екзацербатор	6	7
NPV/PPV	0.6	0.5

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

## XGBoost

Създаден е XGBoost модел за прогнозиране на пациенти с висок риск от екзацербации (група Е по GOLD). Тежестта на отделните предиктори е показана на Фигура 26.



**Фигура 26. Важност на характеристиките в XGBoost модела**

Съкращения: ICS – употреба на инхалаторни кортикостероиди, CC\_ДК – свободна стена на дясна камера (ехохардиография), Plt - тромбоцити, WBC - левкоцити, ФИ – фракция на изтласкване (ехохардиография), postFEV1percent – ФЕО1 (%) след бронходилататор, ФС – фракция на съкращение, Charlson – Charlson индекс на коморбидност, AAT – алфа1 антитрипсин, Packyears – пакетогодини тютюнопушене, SatO2Post – сатурация след 6MWT, SixMWD – изминато разстояние на 6MWT, Wheezing – симптом: свирене в гърдите, TiffneauPre – ФЕО1/ФВК преди бронходилатация, ClinicalPhenotype – клиничен фенотип

Моделът класифицира правилно 65.5% (19/29) от тестовите наблюдения (Таблица 44).

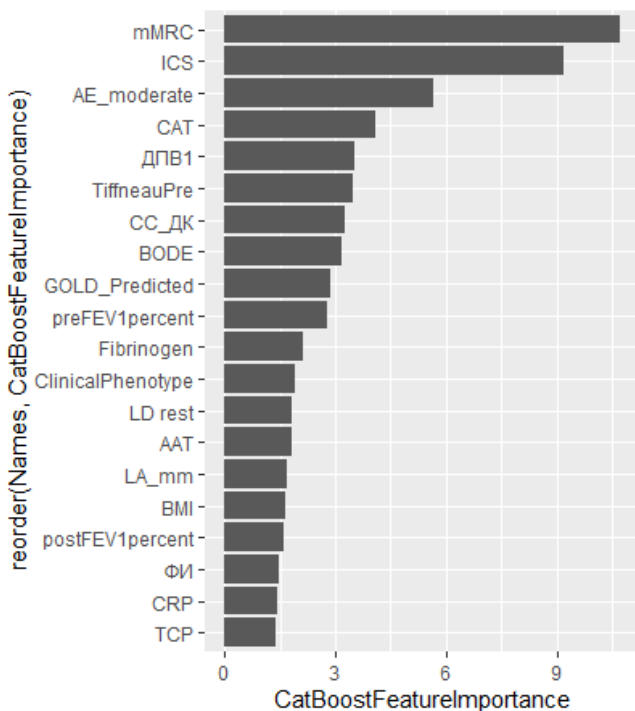
**Таблица 44. Точност на XGBoost модела**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	11	5
Екзацербатор	5	8
NPV/PPV	0.69	0.62

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

### CatBoost

Създаден е CatBoost модел за прогнозиране на пациенти с висок риск от екзацербации (група Е по GOLD). Тежестта на отделните предиктори е показана на Фигура 27.



**Фигура 27. Важност на характеристиките в CatBoost модела**

Съкращения: ICS – употреба на инхалаторни кортикостероиди, AE\_moderate – умерено-тежки екзацербации в предходната година, ДПВ1 – максимален размер на долна празна вена, TiffneauPre – ФЕО1/ФВК преди бронходилатация, CC\_ДК – свободна стена на дясна камера

(ехохардиография), BODE – BODE индекс, GOLD\_Predicted – риск за екзацербация по GOLD (група E), preFEV1percent – ФЕО1 (%) преди бронходилататор, ClinicalPhenotype – клиничен фенотип, LD rest – максимална подвижност на лява диафрагма в покой, ААТ – алфа1 антитрипсин, LA\_mm – размер на ляво предсърдие (ехохардиография), postFEV1percent – ФЕО1 (%) след бронходилататор, ФИ – фракция на изтласкване (ехокардиография), ТСР – телесистолен размер на лява камера (ехокардиография)

Моделът класифицира правилно 72.4% (21/29) от тестовите наблюдения (Таблица 45).

**Таблица 45. Точност на CatBoost модела**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	14	2
Екзацербатор	6	7
NPV/PPV	0.7	0.78

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

### **Ансамблов модел (Ensemble model)**

Ансамбловият модел за прогнозиране на пациенти с висок риск от екзацербации (група E по GOLD) включва трите самостоятелни модела с най-добри резултати. Това са логистична регресия (75.9%), CatBoost (72.4%) и Random Forest (69.0%). Гласуването е осъществено с просто мнозинство – избира се категорията с повече гласове (т.е.  $\geq 2$ ). Моделът класифицира правилно 72.4% (21/29) от тестовите наблюдения (Таблица 46).

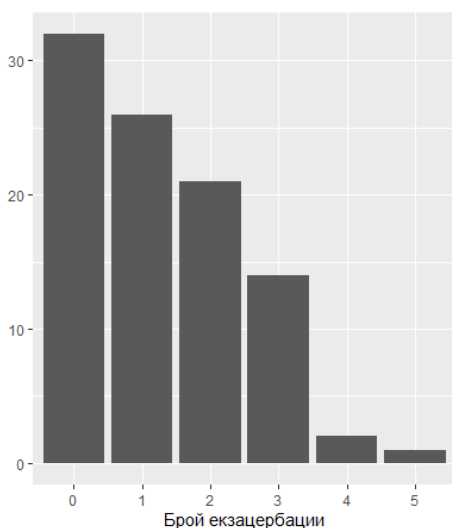
**Таблица 46. Точност на ансамбловия модел**

	Прогноза за неекзацербатор	Прогноза за екзацербатор
Неекзацербатор	13	3
Екзацербатор	5	8
NPV/PPV	0.72	0.73

NPV – Негативна предиктивна стойност, PPV – позитивна предиктивна стойност

#### 4.4. Честота на екзацербации (непрекъсната променлива)

При едногодишно проследяване 32.8% (22/67) от пациентите в тренировъчния набор от данни са без екзацербация, 22.4% (15/67) са с 1 екзацербация, 25.4% (17/67) са с две екзацербации, 16.4% (11/67) са с 3 екзацербации и по 1.5% (1/67) са с четири и пет екзацербации (Фигура 28). Средната стойност на броя на екзацербациите е 1.36, а медианата е 1. С цел възпроизводимост на резултатите ще бъде използван *seed* за случайните числа (*set.seed(100)*)



**Фигура 28. Разпределение на броя на екзацербациите при едногодишно проследяване**

#### Нулев модел

Няма общоприет модел за прогнозиране на общия брой на екзацербациите. Тъй като данните са с неправилно разпределение ( $p < 0.0001$  от Shapiro-Wilk теста), като нулев модел ще бъде използвана медианата на броя на екзацербациите (т.е. нулевият модел прогнозира „1“ екзацербация при всички наблюдения). Коренът от средната квадратична грешка (RMSE) на нулевия модел е 1.10. За сравнение, RMSE при употреба на средната стойност на броя на екзацербациите (т.е. „1.36“) е 1.12 (по-ниските стойности означават по-малка грешка, т.е. по-добър модел).

## Линейна регресия

Проведена е линейна регресия за прогнозиране на общия брой на екзацербациите. Моделът има най-ниска стойност на RMSE (0.970) при 2 броя на предикторите след кръстосано валидиране с 20 подмножества (20-fold).  $R^2$  на модела е 0.63. Резултатите от коефициентите са представени в Таблица 47.

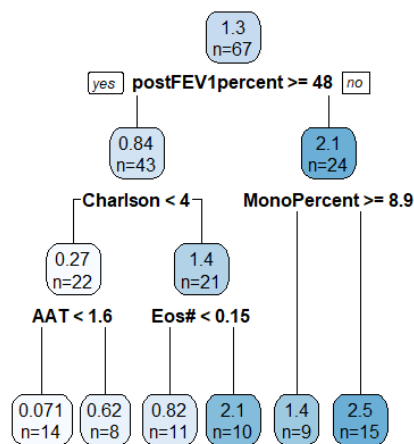
**Таблица 47. Коефициенти на модела с линейна регресия (подредени по абсолютна стойност)**

Параметър	Стойност
Наличие на АСО	1.36
BODE индекс	0.31
Intercept (пресечна точка)	0.23

Забележка: положителните стойности показват повишен риск (повече екзацербации)

## Дърво на решението

Създаден е модел с дърво на решението за прогнозиране на общия брой на екзацербациите (Фигура 29). Стойността на RMSE на модела е 1.329.

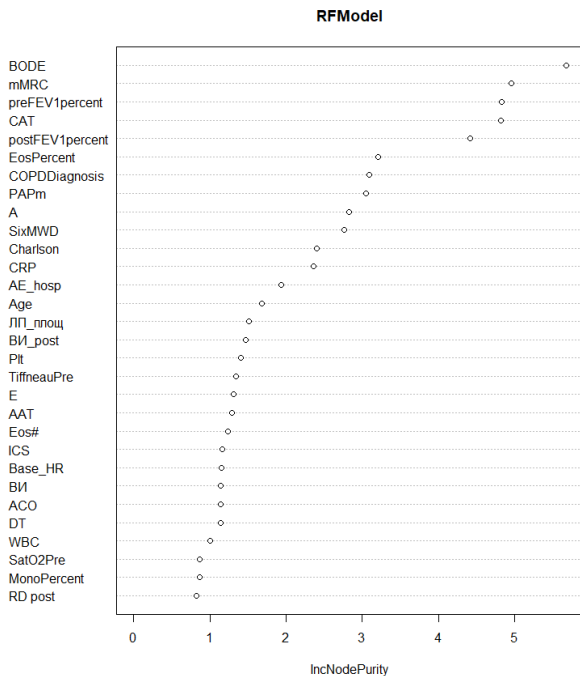


**Фигура 29. Модел на дърво на решението за прогнозиране на общия брой на екзацербациите**

Забележка: postFEV1percent – ФЕО1 (%) след бронходилатация; Charlson – Charlson индекс на коморбидност; MonoPercent – моноцити (%); AAT – ниво на  $\alpha$ -1-антитрипсин (в g/l); Eos# - абсолютен брой еозинофили.

## Random Forest модел

Създаден е Random Forest модел за прогнозиране на общия брой на екзацербациите. Тежестта на отделните предиктори е показана на Фигура 30. Стойността на RMSE на модела е 1.004.

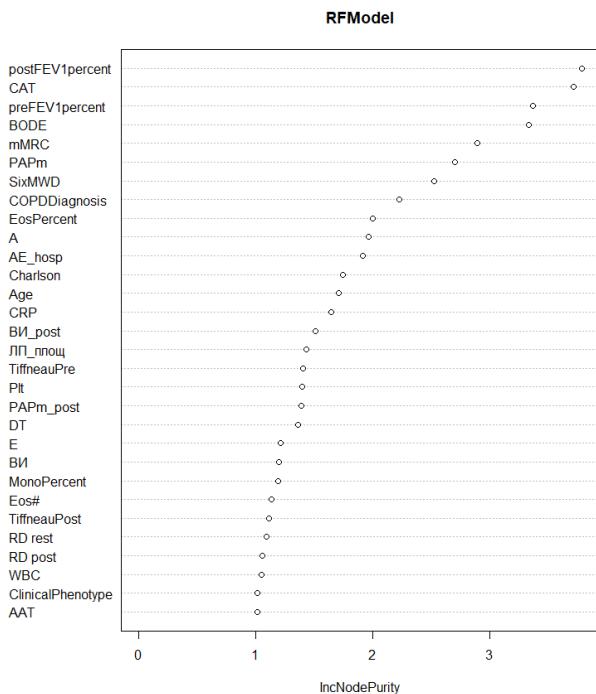


**Фигура 30. Random Forest модел за прогнозиране на общия брой на екзацербациите**

Съкращения: BODE – BODE индекс, preFEV1percent - ФЕО1 (%) преди бронходилататор, postFEV1percent – ФЕО1 (%) след бронходилататор, EosPercent – еозинофили (%), COPDDiagnosis – давност на диагнозата ХОББ, PAPm - средно налягане в а. pulmonalis в покой (ехокардиография), А - скорост на А вълната (ехокардиография), SixMWD – изминато разстояние на 6MWT, Charlson – Charlson индекс на коморбидност, AE\_hosp - тежки екзацербации в предходната година, Age – възраст, ЛП\_площ – площ на ляво предсърдие (ехокардиография), ВИ\_post – време на изтласкване след 6MWT (ехокардиография), Plt - тромбоцити, TiffneauPre – ФЕО1/ФВК преди бронходилатация, E – скорост на E вълната (ехокардиография), AAT – алфа1 антитрипсин, Eos# - еозинофили (абсолютен брой), ICS – употреба на инхалаторни кортикостероиди, Base\_HR – сърдечна честота в покой, ВИ – време на изтласкване

(ехокардиография), АСО – астма-ХОББ припокриване, DT – децелерационно време (ехокардиография), WBC - левкоцити, SatO2Pre – кислородна сатурация преди 6MWT, MonoPercent – моноцити (%), RD post - подвижност на дясна диафрагма след 6MWT.

Търсенето на оптималните параметри показва най-висока точност на модела при  $mtry = 8$  (брой на променливите, които се избират на случаен принцип като кандидати при всяко разделяне) и  $ntree = 2500$  (брой на дърветата). Стойността на RMSE на модела е 0.989. Тежестта на отделните предиктори е показана на Фигура 31.



**Фигура 31. Оптимизиран Random Forest модел за прогнозиране на общия брой на екзацербациите**

Съкращения: postFEV1percent – ФЕО1 (%) след бронходилататор, preFEV1percent - ФЕО1 (%) преди бронходилататор, BODE – BODE индекс, PAPm - средно налягане в а. pulmonalis в покой (ехокардиография), SixMWD – изминато разстояние на 6MWT, COPDDiagnosis – давност на диагнозата ХОББ, EosPercent – еозинофили (%), A - скорост на А вълната (ехокардиография), AE\_hosp - тежки екзацербации в предходната година, Charlson – Charlson индекс на

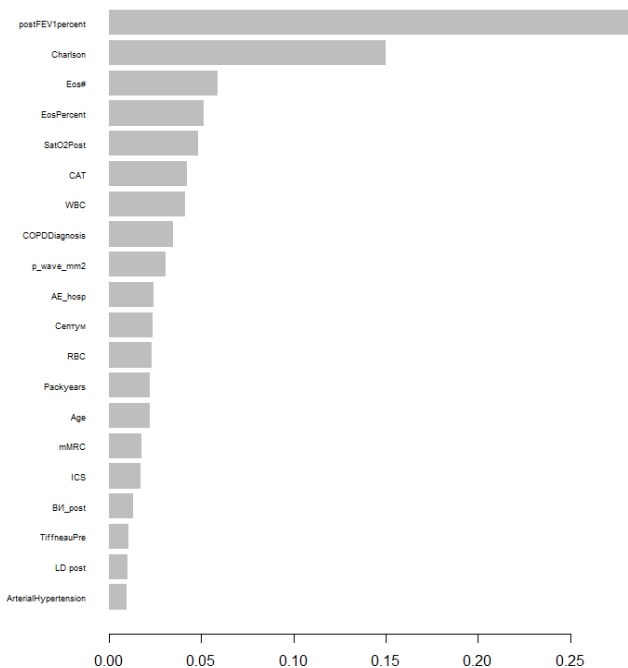
коморбидност, Age – възраст, VI\_post – време на изтласкване след 6MWT, ЛП площ – площ на ляво предсърдие (ехокардиография), TiffneauPre – ФЕО1/ФВК преди бронходилатация, Plt - тромбоцити, PAPm\_post – средно налягане в а. pulmonalis след 6MWT (ехокардиография), DT – децелерационно време (ехокардиография), E – скорост на E вълната (ехокардиография), VI – време на изтласкване (ехокардиография), MonoPercent – моноцити (%), Eos# - еозинофили (абсолютен брой), TiffneauPost – ФЕО1/ФВК след бронходилатация, RD rest - подвижност на дясна диафрагма в покой, RD post - подвижност на дясна диафрагма след 6MWT, WBC - левкоцити, ClinicalPhenotype – клиничен фенотип, AAT – алфа1 антитрипсин.

### **Невронни мрежи (Neural Networks)**

Създаден е модел с невронна мрежа за прогнозиране на общия брой на екзацербациите. Стойността на RMSE на модела е 1.028.

### **XGBoost**

Създаден е XGBoost модел за прогнозиране на общия брой на екзацербациите. Тежестта на отделните предиктори е показана на Фигура 32. Стойността на RMSE на модела е 1.174.

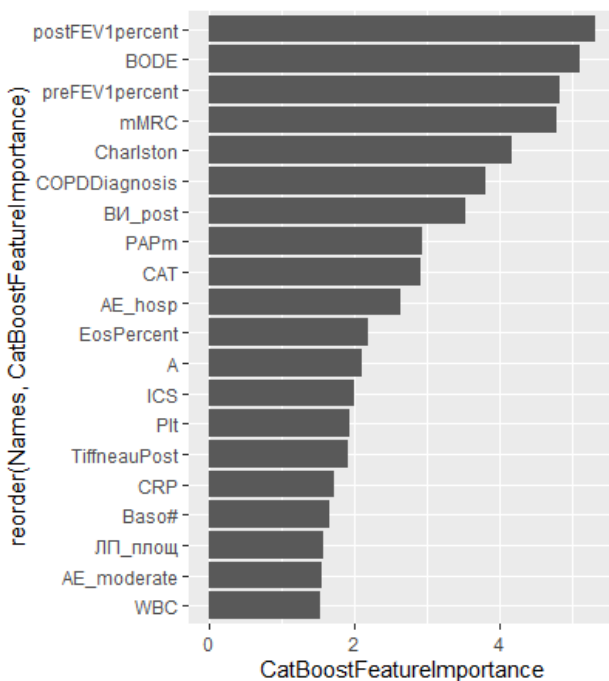


**Фигура 32. Важност на характеристиките в XGBoost модела**

Съкращения: postFEV1percent – ФЕО1 (%) след бронходилататор, Charlson – Charlson индекс на коморбидност, Eos# - еозинофили (абсолютен брой), EosPercent – еозинофили (%), SatO2Post – сатурация след 6MWT, WBC - левкоцити, COPDDiagnosis – давност на диагнозата ХОББ, p\_wave\_mm – височина на р-вълната (ЕКГ), AE\_hosp - тежки екзацербации в предходната година, септум - дебелина на септума (ехокардиография), RBC – еритроцити, Packyears – пакетогодини тютюнопушене, Age – възраст, ICS – лечение с кортикостероиди, BI\_post – време на изтласкване след 6MWT (ехокардиография), TiffneauPre – ФЕО1/ФВК преди бронходилатация, LD post - подвижност на лява диафрагма след 6MWT, ArterialHypertension – анамнеза за артериална хипертония.

### CatBoost

Създаден е CatBoost модел за прогнозиране на общия брой на екзацербациите. Тежестта на отделните предиктори е показана на Фигура 33. RMSE на модела е 1.010.



**Фигура 33. Важност на характеристиките в CatBoost модела**

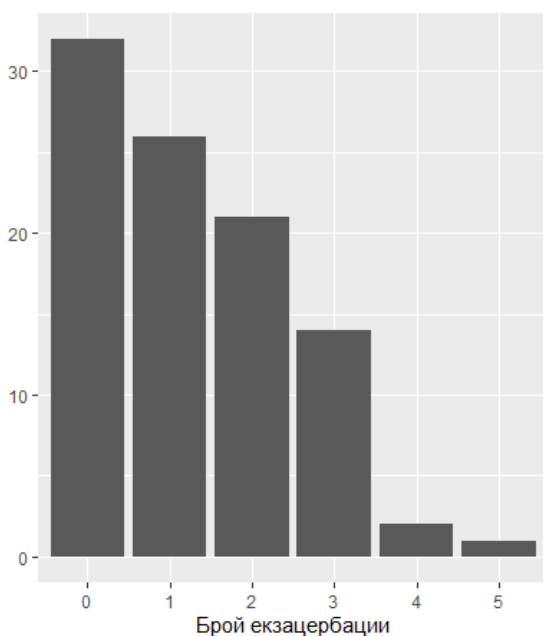
Съкращения: postFEV1percent – ФЕО1 (%) след бронходилататор, BODE – BODE индекс, preFEV1percent – ФЕО1 (%) преди бронходилататор, Charlston – Charlson индекс на коморбидност, COPDDiagnosis – давност на диагнозата ХОББ, BI\_post – време на изтласкване след 6MWT (ехокардиография), PAPm - средно налягане в а. pulmonalis в покой (ехокардиография), AE\_hosp - тежки екзацербации в предходната година, EosPercent – еозинофили (%), A - скорост на А вълната (ехокардиография), ICS – лечение с кортикостероиди, Plt – тромбоцити, TiffneauPost – ФЕО1/ФВК след бронходилатация, Baso# - базофили (абсолютен брой), ЛП\_плоч – площ на ляво предсърдие (ехокардиография), AE\_moderate – умерено-тежки екзацербации в предходната година, WBC – левкоцити.

### Ансамблов модел (Ensemble model)

Ансамбловият модел за прогнозиране на общия брой на екзацербациите (непрекъсната променлива) включва трите самостоятелни модела с най-добри резултати – линейна регресия (0.970), Random Forest (0.989) и CatBoost (1.010). Гласуването е осъществено чрез определяне на средната стойност на трите модела. RMSE на модела е 0.959.

#### 4.5. Честота на екзацербации (категорийна променлива)

При едногодишно проследяване 32.8% (22/67) от пациентите в тренировъчния набор от данни са без екзацербация, 22.4% (15/67) са с 1 екзацербация, 25.4% (17/67) са с две екзацербации, 16.4% (11/67) са с 3 екзацербации и по 1.5% (1/67) са с четири и пет екзацербации (Фигура 34). Средната стойност на броя на екзацербациите е 1.36, а медианата е 1. С цел възпроизводимост на резултатите ще бъде използван *seed* за случайните числа (*set.seed(100)*)



**Фигура 34. Разпределение на броя на екзацербациите при едногодишно проследяване**

#### Нулев модел

Няма общоприет модел за прогнозиране на броя на екзацербациите (категорийна променлива). Тъй като данните са с неправилно разпределение ( $p < 0.0001$  от Shapiro-Wilk теста) като

нулев модел ще бъде използвана най-често срещаната категория (модата) на броя на екзацербациите (т.е. нулевият модел прогнозира „0“ екзацербации при всички наблюдения). Моделът класифицира правилно 34.5% (10/29) от тестовите наблюдения (Таблица 48).

**Таблица 48. Точност на нулевия модел**

Действител ен брой АЕ	Прогно за за „0“ АЕ		Прогно за за „1“ АЕ		Прогно за за „2“ АЕ		Прогно за за „3“ АЕ		Прогно за за „4“ АЕ		Прогно за за „5“ АЕ	
	<b>0</b>	<b>10</b>	0	0	0	0	0	0	0	0	0	0
<b>1</b>	8	<b>0</b>	0	0	0	0	0	0	0	0	0	0
<b>2</b>	6	0	<b>0</b>	0	0	0	0	0	0	0	0	0
<b>3</b>	4	0	0	<b>0</b>	0	0	<b>0</b>	0	0	0	0	0
<b>4</b>	1	0	0	0	0	0	0	<b>0</b>	0	0	0	0

Съкращения: АЕ – екзацербация/и

### Логистична регресия

Проведена е логистична регресия за прогнозиране на броя на екзацербациите (категорийна променлива) с помощта на функцията *multinom* от пакета *nnet* в RStudio, за да се създаде многокатегориен логистичен регресионен модел. Резултатите от коефициентите са представени в Таблица 49. Моделът класифицира правилно 31.0% (9/29) от тестовите наблюдения (Таблица 50).

**Таблица 49. Коефициенти на модела с логистична регресия (подредени по абсолютна стойност)**

Параметър	Стойност
Наличие на АСО	1.36
BODE индекс	0.308
Intercept (пресечна точка)	0.225

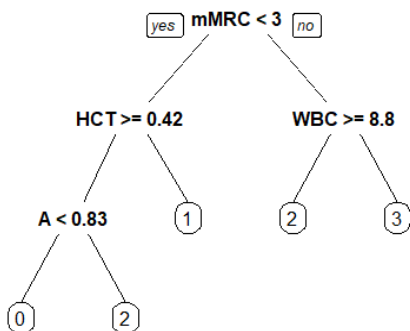
**Таблица 50. Точност на модела с логистична регресия**

Действител ен брой АЕ	Прогно за за „0“ АЕ		Прогно за за „1“ АЕ		Прогно за за „2“ АЕ		Прогно за за „3“ АЕ		Прогно за за „4“ АЕ		Прогно за за „5“ АЕ	
	<b>0</b>	<b>3</b>	0	0	5	1	1	0	0	0	0	0
<b>1</b>	1	<b>2</b>	0	4	0	0	0	0	0	0	1	0
<b>2</b>	1	1	1	<b>3</b>	0	1	0	0	0	0	0	0
<b>3</b>	1	0	0	2	0	<b>1</b>	0	0	0	0	0	0
<b>4</b>	0	0	0	1	0	0	<b>0</b>	0	0	0	0	0

Съкращения: АЕ – екзацербация/и

### Дърво на решението

Създаден е модел с дърво на решението за прогнозиране на броя на екзацербациите (категорийна променлива) (Фигура 35). Моделът класифицира правилно 31.0% (9/29) от тестовите наблюдения (Таблица 51).



**Фигура 35. Модел на дърво на решението за прогнозиране на броя на екзацербациите (категорийна променлива)**

Забележка: А = скорост на А вълната (ехокардиография), HCT – хематокрит, WBC - левкоцити

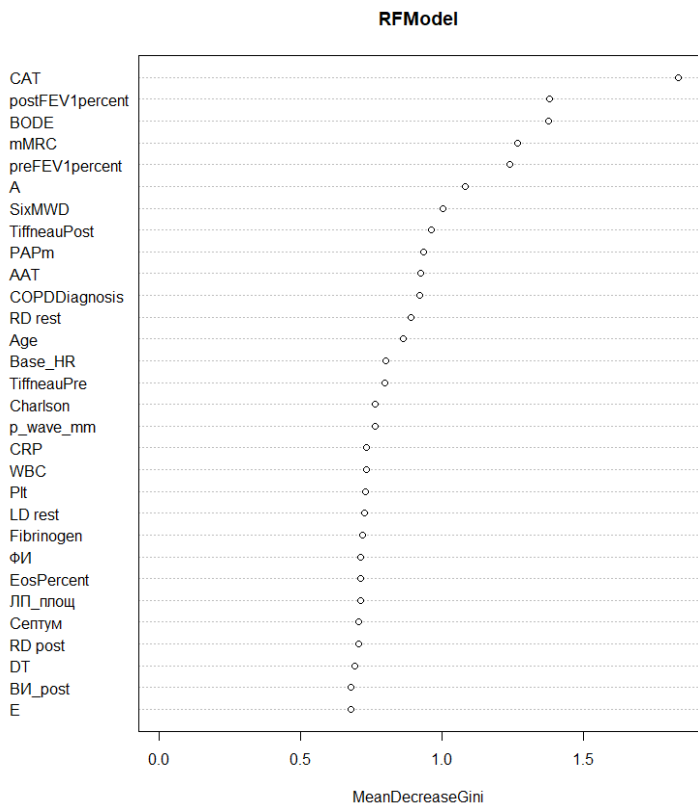
**Таблица 51. Точност на модела с дърво на решението**

Действителен брой АЕ	Прогно за за		Прогно за за		Прогно за за		Прогно за за		Прогно за за	
	„0“ АЕ	АЕ	„1“ АЕ	АЕ	„2“ АЕ	АЕ	„3“ АЕ	АЕ	„4“ АЕ	АЕ
<b>0</b>	<b>4</b>	4	2	0	0	0	0	0	0	0
<b>1</b>	3	<b>1</b>	1	3	0	0	0	0	0	0
<b>2</b>	3	0	<b>2</b>	1	0	0	0	0	0	0
<b>3</b>	1	0	1	<b>2</b>	0	0	0	0	0	0
<b>4</b>	0	0	1	0	<b>0</b>	0	0	0	0	0

Съкращения: АЕ – екзацербация/и

### Random Forest модел

Създаден е Random Forest модел за прогнозиране на броя на екзацербациите (категорийна променлива). Тежестта на отделните предиктори е показана на Фигура 36. Моделът класифицира правилно 41.4% (12/29) от тестовите наблюдения (Таблица 52).



**Фигура 36. Random Forest модел за прогнозиране на броя на екзацербациите (категорийна променлива)**

Съкращения: postFEV1percent – ФЕО1 (%) след бронходилататор, BODE – BODE индекс, preFEV1percent - ФЕО1 (%) преди бронходилататор, А - скорост на А вълната (ехокардиография), SixMWD – изминато разстояние на 6MWT, TiffneauPost – ФЕО1/ФВК след бронходилатация, PAPm - средно налягане в а. pulmonalis в покой (ехокардиография), ААТ – алфа1 антитрипсин, COPDDiagnosis – давност на диагнозата ХОББ, RD rest – подвижност на дясна диафрагма в покой, Age – възраст, Base\_HR – сърдечна честота в покой, TiffneauPre – ФЕО1/ФВК преди бронходилатация, Charlson – Charlson индекс на коморбидност, p\_wave\_mm – височина на р-вълната (ЕКГ), WBC - левкоцити, Plt - тромбоцити, LD rest - подвижност на лява диафрагма в покой, ФИ - фракцията на изтласкване (ехокардиография), EosPercent – еозинофили (%), ЛП\_площ – площ на ляво предсърдие (ехокардиография), септум - дебелина на септума (ехокардиография), RD post - подвижност на дясна диафрагма след 6MWT, DT – децелерационно време (ехокардиография),

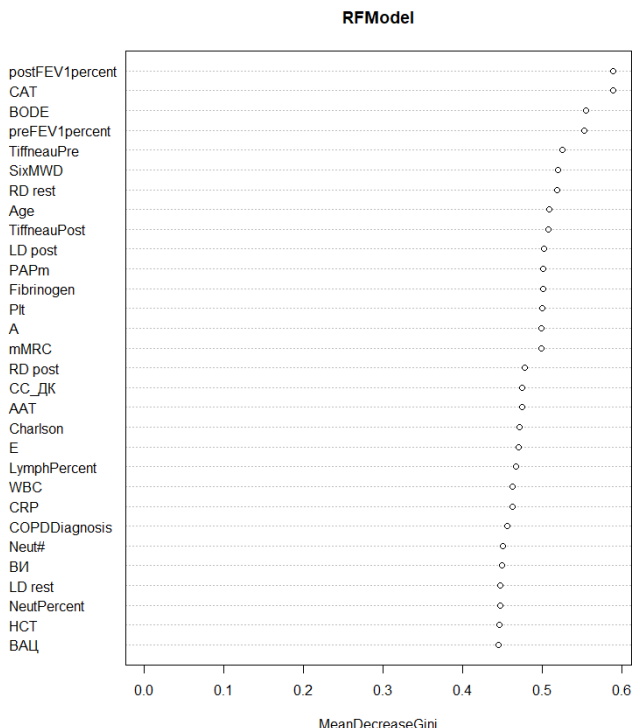
VI post – време на изтласкване след 6MWT (ехокардиография), E – скорост на E вълната (ехокардиография).

**Таблица 52. Точност на Random Forest модела**

Действителен брой АЕ	Прогно за за		Прогно за за		Прогно за за		Прогно за за		Прогно за за	
	„0“ АЕ	„1“ АЕ	„2“ АЕ	„3“ АЕ	„4“ АЕ	„5“ АЕ	„0“ АЕ	„1“ АЕ	„2“ АЕ	„3“ АЕ
<b>0</b>	<b>8</b>	1	1	0	0	0	0	0	0	0
<b>1</b>	2	<b>3</b>	3	0	0	0	0	0	0	0
<b>2</b>	2	4	<b>0</b>	0	0	0	0	0	0	0
<b>3</b>	1	2	0	<b>1</b>	0	0	0	0	0	0
<b>4</b>	0	0	1	0	<b>0</b>	0	0	0	0	0

Съкращения: АЕ – екзацербация/и

Търсенето на оптималните параметри показва най-висока точност на модела при  $mtry = 1$  (брой на променливите, които се избират на случаен принцип като кандидати при всяко разделяне) и  $ntree = 3000$  (брой на дърветата). Тежестта на отделните предиктори е показана на Фигура 37. Моделът класифицира правилно 44.8% (13/29) от наблюденията (Таблица 53).



**Фигура 37. Оптимизиран Random Forest модел за прогнозиране на броя на екзацербациите (категорийна променлива)**

Съкращения: postFEV1percent – ФЕО1 (%) след бронходилататор, BODE – BODE индекс, preFEV1percent - ФЕО1 (%) преди бронходилататор, TiffneauPre – ФЕО1/ФВК преди бронходилатация, SixMWD – изменато разстояние на 6MWT, RD rest – подвижност на дясна диафрагма в покой, Age – възраст, TiffneauPost – ФЕО1/ФВК след бронходилатация, LD post - подвижност на лява диафрагма след 6MWT, PAPm - средно налягане в а. pulmonalis в покой, Plt - тромбоцити, A - скорост на А вълната (ехокардиография), RD post - подвижност на дясна диафрагма след 6MWT, CC\_ДК – свободна стена на дясна камера, AAT – алфа1 антитрипсин, Charlson – Charlson индекс на коморбидност, E – скорост на Е вълната (ехокардиография), LymphPercent – лимфоцити (%), WBC - левкоцити, COPDDiagnosis – давност на диагнозата ХОББ, Neut# - неутрофили (абсолютен брой), BI – време на изтласкване, LD rest - подвижност на лява диафрагма в покой, NeutPercent – неутрофили (%), HCT – хематокрит, BAЦ – време на акцелерация.

**Таблица 53. Точност на оптимизирания Random Forest модел**

Действител	Прогно	Прогно	Прогно	Прогно	Прогно	Прогно
------------	--------	--------	--------	--------	--------	--------

ен брой АЕ	за за „0“ АЕ	за за „1“ АЕ	за за „2“ АЕ	за за „3“ АЕ	за за „4“ АЕ	за за „5“ АЕ
<b>0</b>	<b>9</b>	1	0	0	0	0
<b>1</b>	3	<b>4</b>	1	0	0	0
<b>2</b>	4	2	<b>0</b>	0	0	0
<b>3</b>	2	2	0	<b>0</b>	0	0
<b>4</b>	0	1	0	0	<b>0</b>	0

Съкращения: АЕ – екзацербация/и

### Невронни мрежи (Neural Networks)

Създаден е модел с невронна мрежа за прогнозиране на броя на екзацербациите (категорийна променлива). Моделът класифицира правилно 34.5% (10/29) от наблюденията (Таблица 54).

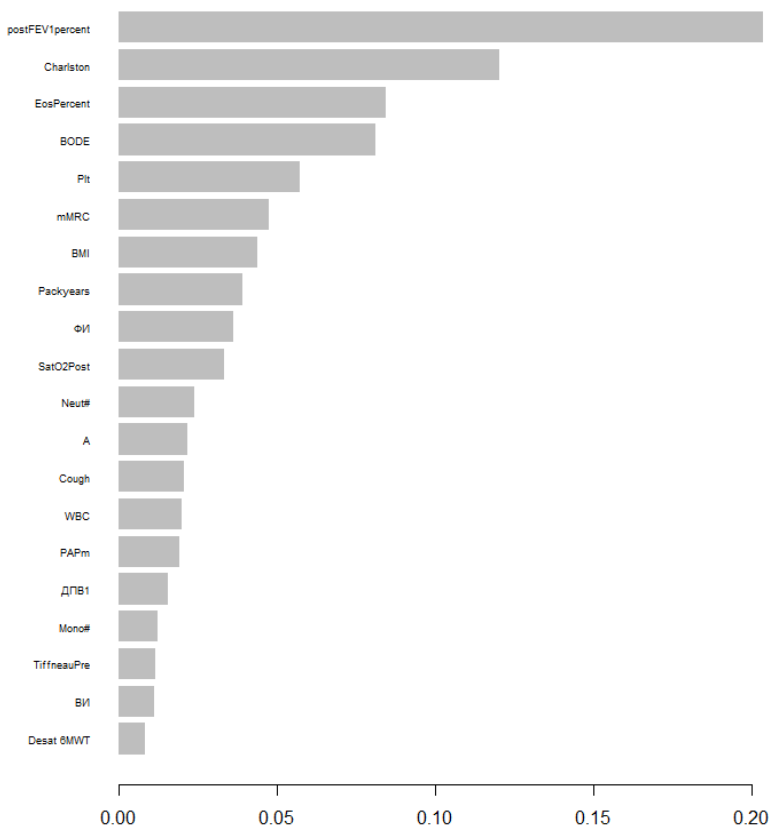
**Таблица 54. Точност на модела с невронна мрежа**

Действител ен брой АЕ	Прогно за за „0“ АЕ	Прогно за за „1“ АЕ	Прогно за за „2“ АЕ	Прогно за за „3“ АЕ	Прогно за за „4“ АЕ	Прогно за за „5“ АЕ
<b>0</b>	<b>6</b>	1	2	1	0	0
<b>1</b>	2	<b>2</b>	2	2	0	0
<b>2</b>	1	2	<b>2</b>	1	0	0
<b>3</b>	2	1	1	<b>0</b>	0	0
<b>4</b>	0	0	1	0	<b>0</b>	0

Съкращения: АЕ – екзацербация/и

### XGBoost

Създаден е XGBoost модел за прогнозиране на броя на екзацербациите (категорийна променлива). Тежестта на отделните предиктори е показана на Фигура 38. Моделът класифицира правилно 37.9% (11/29) от наблюденията (Таблица 55).



**Фигура 38. XGBoost модел за прогнозиране на броя на екзацербациите (категорийна променлива)**

Съкращения: postFEV1percent – ФЕО1 (%) след бронходилататор, Charlson – Charlson индекс на коморбидност, EosPercent – еозинофили (%), BODE – BODE индекс, Plt - тромбоцити, Packyears – пакетогодини тютюнопушене, ФИ - фракцията на изтласкване (ехокардиография), SatO2Post – сатурация след 6MWT, Neut# - неутрофили (абсолютен брой), A - скорост на А вълната (ехокардиография), Cough – симптом: кашлица, WBC - левкоцити, PAPm - средно налягане в а. pulmonalis в покой (ехокардиография), ДПВ1 – максимален размер на долна празна вена, Mono# - моноцити (абсолютен брой), TiffneauPre – ФЕО1/ФВК преди бронходилатация, VI – време на изтласкване (ехокардиография), Desat 6MWT – десатурационен индекс след 6MWT.

**Таблица 55. Точност на XGBoost модела**

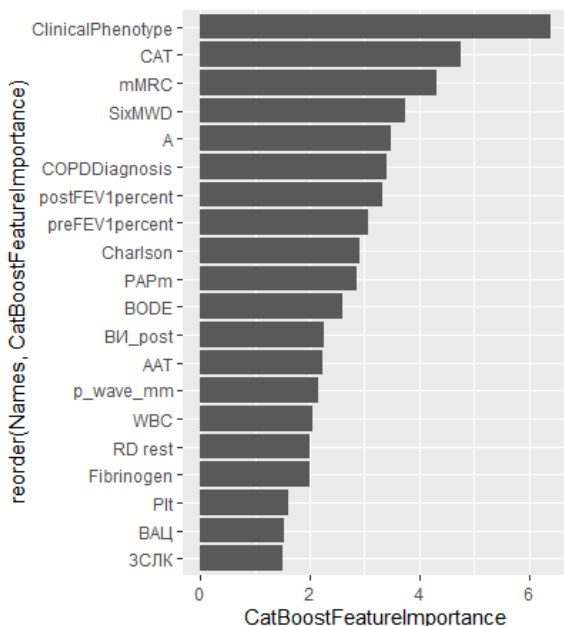
Действител	Прогно	Прогно	Прогно	Прогно	Прогно	Прогно
------------	--------	--------	--------	--------	--------	--------

ен брой АЕ	за за	за за	за за	за за	за за	за за
	„0“ АЕ	„1“ АЕ	„2“ АЕ	„3“ АЕ	„4“ АЕ	„5“ АЕ
<b>0</b>	<b>6</b>	1	3	0	0	0
<b>1</b>	1	<b>3</b>	4	0	0	0
<b>2</b>	1	3	<b>2</b>	0	0	0
<b>3</b>	0	3	1	<b>0</b>	0	0
<b>4</b>	0	0	1	0	<b>0</b>	0

Съкращения: АЕ – екзацербация/и

## CatBoost

Създаден е CatBoost модел за прогнозиране на броя на екзацербациите (категорийна променлива). Тежестта на отделните предиктори е показана на Фигура 39. Моделът класифицира правилно 48.3% (14/29) от наблюденията (Таблица 56).



**Фигура 39. CatBoost модел за прогнозиране на броя на екзацербациите (категорийна променлива)**

Съкращения: ClinicalPhenotype – клиничен фенотип, SixMWD – изминато разстояние на 6MWT, A – скорост на A вълната (ехокардиография), COPDDiagnosis – давност на диагнозата ХОББ, postFEV1percent – ФЕО1 (%) след бронходилататор, preFEV1percent - ФЕО1 (%) преди бронходилататор, Charlson – Charlson индекс на коморбидност, PAPm -

средно налягане в а. pulmonalis в покой (ехокардиография), BODE – BODE индекс, VI\_post – време на изтласкване след 6MWT (ехокардиография), AAT – алфа1 антитрипсин, p\_wave\_mm – височина на р-вълната (ЕКГ), WBC - левкоцити, RD rest - подвижност на дясна диафрагма в покой, Plt - тромбоцити, ВАЦ – време на акцелерация (ехокардиография), ЗСЛК – задна стена на лява камера (ехокардиография).

**Таблица 56. Точност на CatBoost модела**

Действител ен брой АЕ	Прогно за за		Прогно за за		Прогно за за		Прогно за за		Прогно за за	
	„0“ АЕ	АЕ	„1“ АЕ	АЕ	„2“ АЕ	АЕ	„3“ АЕ	АЕ	„4“ АЕ	АЕ
<b>0</b>	<b>9</b>		0		1		0		0	
<b>1</b>	2		<b>3</b>		3		0		0	
<b>2</b>	2		2		<b>2</b>		0		0	
<b>3</b>	1		2		1		<b>0</b>		0	
<b>4</b>	0		0		1		0		<b>0</b>	

Съкращения: АЕ – екзацербация/и

### Ансамблов модел (Ensemble model)

Ансамбловият модел за прогнозиране на броя на екзацербациите (категорийна променлива) включва трите самостоятелни модела с най-добри резултати - CatBoost (48.3%), Random Forest (44.8%) и XgBoost (34.5%). Гласуването е осъществено с просто мнозинство – избира се категорията с повече гласове (т.е.  $\geq 2$ ). В случай на равенство (например трите модела гласуват за три различни категории) се избира категорията на най-добрия модел. Моделът класифицира правилно 41.4% (12/29) от тестовите наблюдения (Таблица 57).

**Таблица 57. Точност на ансамбловия модел**

Действител ен брой АЕ	Прогно за за		Прогно за за		Прогно за за		Прогно за за		Прогно за за	
	„0“ АЕ	АЕ	„1“ АЕ	АЕ	„2“ АЕ	АЕ	„3“ АЕ	АЕ	„4“ АЕ	АЕ
<b>0</b>	<b>8</b>		1		1		0		0	
<b>1</b>	1		<b>2</b>		5		0		0	
<b>2</b>	1		3		<b>2</b>		0		0	
<b>3</b>	0		3		1		<b>0</b>		0	
<b>4</b>	0		0		1		0		<b>0</b>	

Съкращения: АЕ – екзацербация/и

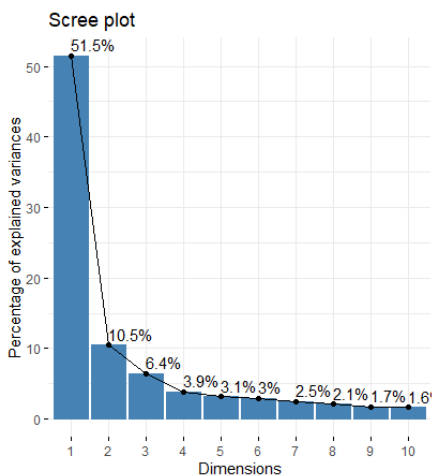
## 4.6. Фенотипизиране

### Метод на главните елементи (principal components analysis, PCA)

Резултатите от проведения анализ чрез метода на главните елементи са показани на Таблица 58. В ред "Кумулативна пропорция" първият главен компонент обяснява 51.5% от общата дисперсия. Това означава, че половината от данните могат да бъдат представени само от първия главен компонент (Фигура 40). Вторият компонент обяснява 10.5% от общата дисперсия. Кумулативният дял на първите 3 компонента обяснява над 2/3 от общата дисперсия.

**Таблица 58. Важност на първите пет компонента от PCA**

Компонент	1	2	3	4	5
Стандартно отклонение	1.50	0.68	0.53	0.41	0.37
Дял от дисперсията	0.52	0.11	0.06	0.04	0.03
Кумулативна пропорция	0.52	0.62	0.68	0.72	0.75



**Фигура 40. Скала на компонентите**

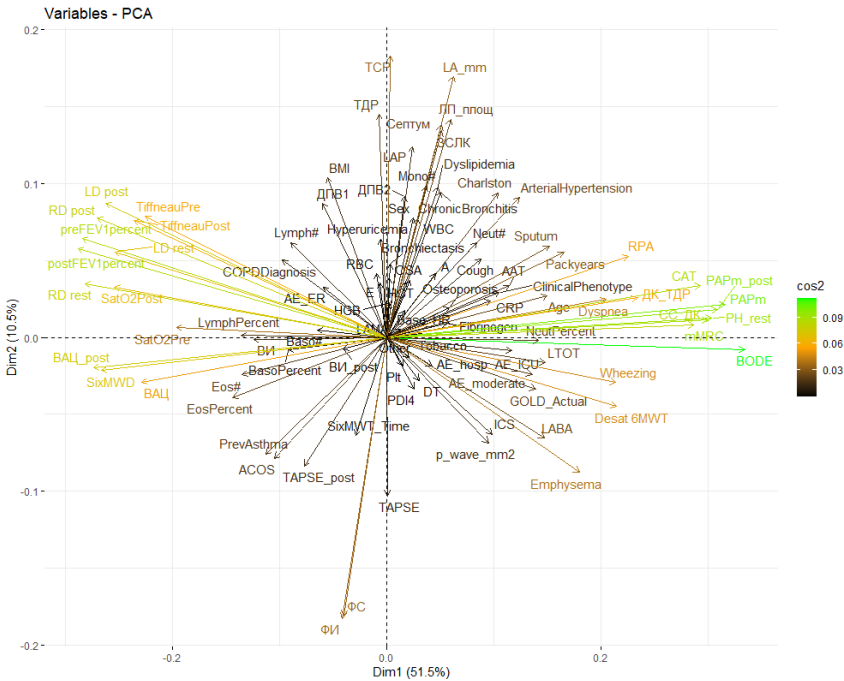
На Таблица 59 са показани, елементите с най-голямо значение в първите три компонента. За първият компонент с най-голямо значение е BODE индексът, PAPm в покой и след натоварване и качеството на живот. За вторият компонент с най-голямо значение са показатели от ехокардиографията, а за третия – ехокардиографски и лабораторни показатели.

**Таблица 59. Матрица на първите три главни компонента**

Компонент		1	2	3
Елемент (стойност)	1	BODE (0.22)	ТСР (0.27)	ВИ (-0.24)
Елемент (стойност)	2	PAPm (0.21)	ФС (-0.27)	ВИ_post (-0.24)
Елемент (стойност)	3	PAPm_post (0.21)	ФИ (-0.27)	NeutPercent (0.23)
Елемент (стойност)	4	mMRC (0.20)	LA_mm (0.25)	Е (-0.20)
Елемент (стойност)	5	CAT (0.20)	ЛП_плоч (0.21)	EosPercent (-0.20)

Съкращения: BODE – BODE индекс, PAPm - средно налягане в а. pulmonalis в покой (ехокардиография), PAPm\_post - средно налягане в а. pulmonalis след 6MWT (ехокардиография), ТСР - телесистолен размер на лява камера (ехокардиография), ФС – фракция на съкращение (ехокардиография), ФИ - фракция на изтласкване (ехокардиография), LA\_mm - размер на ляво предсърдие (ехокардиография), ЛП\_плоч – площ на ляво предсърдие (ехокардиография), ВИ - време на изтласкване в покой (ехокардиография), ВИ\_post – време на изтласкване след 6MWT (ехокардиография), NeutPercent - неутрофили (%), Е - скорост на Е вълната (ехокардиография), EosPercent – еозинофили (%)

Влиянието на характеристиките за всеки от главните компоненти е показано на Фигура 41. По-високите стойности на *cos2* (косинуса на втора степен), отбелязани със зелено и оранжево, показват по-висока важност на променливата за компонентите.



**Фигура 41. Графика на променливите в зависимост от главните компоненти**

Означенията са като в останалите фигури.

Оси: Двете оси, обозначени като "Dim1" и "Dim2", представляват съответно първия и втория главен компонент. Процентите (например 51.5% за Dim1) показват каква част от общата дисперсия в набора от данни е обхваната от всеки компонент.

Вектори (стрелки): Представяват оригиналните променливи в набора от данни. Посоката и дължината на стрелката показват как всяка променлива допринася за двата основни компонента. Променливите, които са близо една до друга и в една и съща посока, имат сходни профили в наблюденията.

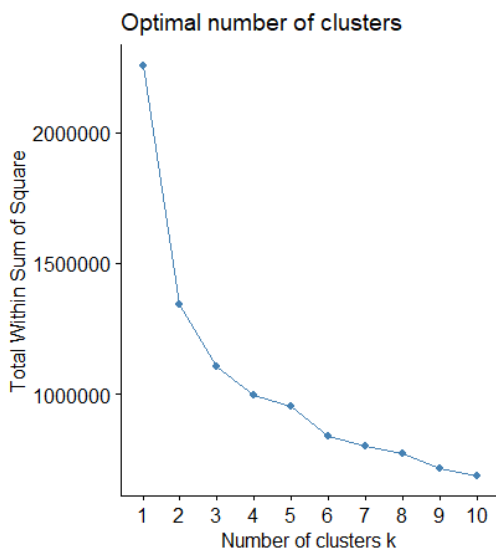
Цветен градиент: Цветният градиент вдясно е скала, показваща приноса или тежестта на всяка променлива.

Означения: Всяко означение съответства на променлива от оригиналния набор от данни.

### Метод на К-средните (K-Means Clustering)

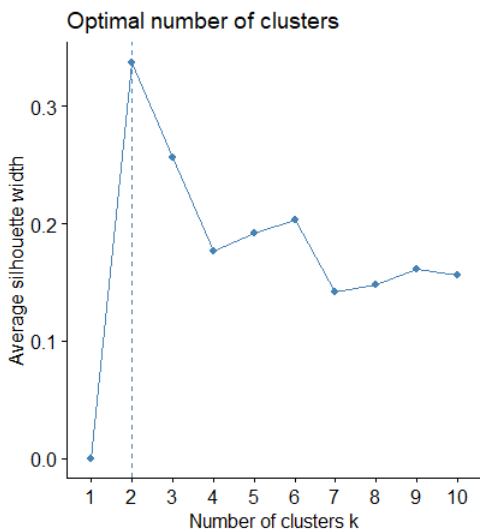
Методът на К-средните започва с определяне на броя на клъстерите, на които ще бъде разделена извадката. На Фигура 42 са показани резултатите от *elbow* метода за определяне на оптималния

брой кълстери. Според този метод извадката може да бъде разделена на два, три или четири кълстера. Разделянето на повече от 4 кълстера е свързано с намалена степен на информативност.



**Фигура 42. Определяне на оптимален брой кълстери чрез *elbow* метода**

На Фигура 43 са показани резултатите от метода на средния силует за определяне на оптималния брой кълстери. Според този метод оптималният брой кълстери е 2.



**Фигура 43. Определяне на оптимален брой кълстери чрез метода на средния силует**

Разделянето на извадката на два кълстера обособява следните групи (Таблица 60):

- в кълстер 1 попадат предимно пациенти с предходна диагноза бронхиална астма и АСО, с по-леко заболяване (ФЕО1) и по-малко симптоми, по-добро качество на живот (САТ и mMRC) и по-добра диафрагмална подвижност;
- в кълстер 2 попадат пациенти с авансирала ХОББ (по-нисък ФЕО1) с наличие както на емфиземен, така и на бронхитен фенотип, значими коморбидности (OSA, артериална хипертония, дислипидемия) и по-нисък физически капацитет (6MWT).

Трябва да се отбележи, че разликата в ВМІ между двата кълстера е 0.1 (26.3 срещу 26.2).

**Таблица 60. Средни стойности на параметрите в отделните кълстери (анализ с 2 кълстера)**

Характеристика	Кълстер 1	Кълстер 2	Средна стойност за извадката
Пол, мъже	60.4%	65.1%	62.5%
Възраст, години	63.2	67.4	65.1
Тютюнопушене,	23.7	34.8	28.7

пакетогодини			
ВМІ	26.3	26.2	26.3
Предходна диагноза bronхиална астма	17.0%	4.7%	11.5%
АСО	18.9%	4.7%	12.5%
Емфизем (фенотип)	47.2%	69.8%	57.3%
Хроничен бронхит (фенотип)	56.6%	72.1%	63.5%
OSA	3.8%	9.3%	6.3%
Артериална хипертония	64.2%	86.0%	74.0%
Дислипидемия	9.4%	27.9%	17.7%
ФЕО1 след бронходилататор	62.2%	47.9%	55.8%
Симптом: свирене в гърдите	60.4%	86.0%	71.9%
Симптом: експекторация	62.3%	83.7%	71.9%
Симптом: задух	71.7%	95.3%	82.3%
Умерено-тежки екзацербации в предходната година	0.58	0.84	0.70
Тежки екзацербации в предходната година	1.01	1.14	1.07
mMRC	1.51	2.74	2.1
CAT	10.7	21.7	15.6
Еозинофили, абсолютен брой	0.25	0.20	0.23
CRP (mg/dL)	3.59	5.42	4.41
Изминато разстояние на 6MWT, m	457.5	267.4	372.3
Десатурация след 6MWT, %	3.6	5.7	4.5
PAPm, mmHg	19.0	26.5	22.4
Наличие на пулмонална хипертония	30.2%	79.1%	52.1%
Подвижност на дясна диафрагма в покой, mm	52.6	42.5	48.1
Общ брой екзацербации след 1 година	0.94	1.70	1.28

При сравняване на резултатите с йерархичната клъстеризация се наблюдават по-ниски разлики (по-лоши резултати) между отделните групи. По-висока разлика се наблюдава единствено при средните стойности на CRP (1.83 срещу 1.35).

Разделянето на извадката на три клъстера обособява следните групи (Таблица 61):

- в кълъстер 1 попадат пациенти с по-добро качество на живот и характеристики на заболяването;
- в кълъстер 2 попадат пациенти с по-висока честота на АСО и междинни характеристики между кълъстер 1 и кълъстер 3.
- в кълъстер 3 попадат пациенти с по-неблагоприятни характеристики на заболяването и влошено качество на живот

В този случай алгоритъмът показва възможност за стратифициране въз основа на тежестта на заболяването. Трябва да се отбележи, че обособените кълъстери се различават по отношение на честотата на екзацербациите при проследяване (0.81 срещу 1.13 срещу 1.84).

**Таблица 61. Средни стойности на параметрите в отделните кълъстери (анализ с 3 кълъстера)**

Характеристика	Кълъстер 1	Кълъстер 2	Кълъстер 3	Средна стойност за извадката
Пол, мъже	69.4%	47.8%	64.9	62.5%
Възраст, години	64.1	62.7	67.6	65.1
Тютюнопушене, пакетогодини	23.9	27.2	34.3	28.7
ВМІ	26.6	25.8	26.1	26.3
Предходна диагноза бронхиална астма	11.1%	26.1%	5.4%	11.5%
АСО	11.1%	26.1%	5.4%	12.5%
Емфизем (фенотип)	52.7%	39.1%	73.0%	57.3%
Хроничен бронхит (фенотип)	63.9%	56.5%	67.6%	63.5%
OSA	5.6%	4.3%	8.1%	6.3%
Артериална хипертония	63.9%	69.6%	86.5%	74.0%
Дислипидемия	8.3%	26.1%	21.6%	17.7%
ФЕО1 след бронходилататор	64.5%	58.0%	45.9%	55.8%
Симптом: свирене в гърдите	50.0%	78.3%	89.2%	71.9%
Симптом: експекторация	58.3%	73.9%	83.8%	71.9%
Симптом: задух	61.1%	95.7%	94.6%	82.3%
Умерено-тежки екзацербации в	0.50	0.70	0.89	0.70

предходната година				
Тежки екзацербации в предходната година	0.89	0.91	1.11	1.07
mMRC	1.25	2.17	2.78	2.1
CAT	9.5	13.5	22.9	15.6
Еозинофили, абсолютен брой	0.24	0.24	0.21	0.23
CRP (mg/dL)	3.26	4.24	5.65	4.41
Изминато разстояние на 6MWT, m	478.2	394.3	255.7	372.3
Десатурация след 6MWT, %	3.8	3.7	5.8	4.5
PAPm, mmHg	18.5	20.7	27.1	22.4
Наличие на пулмонална хипертония	25.0%	43.5%	83.8%	52.1%
Подвижност на дясна диафрагма в покой, mm	53.7	48.6	42.3	48.1
Общ брой екзацербации след 1 година	0.81	1.13	1.84	1.28

Разделянето на извадката на четири клъстера обособява следните групи (Таблица 62):

- в клъстер 1 попадат предимно мъже с добро качество на живот (CAT и mMRC), по-добра белодробна функция (ФЕО1) и съхранена физическа активност (6MWT);
- в клъстер 2 попадат пациенти с предходна диагноза бронхиална астма и АСО, които по показатели заемат междинни стойности между клъстер 1 и клъстер 3;
- в клъстер 3 попадат предимно пациенти с авансирала ХОББ (по-нисък ФЕО1) с наличие на емфиземен фенотип, пулмонална хипертония, намалена диафрагмална подвижност и по-нисък физически капацитет (6MWT);
- в клъстер 4 попадат пациенти които по показатели заемат междинни стойности между клъстер 1 и клъстер 3. Отличителна характеристика на този клъстер е ниската честота на дислипидемия.

Трябва да се отбележи значимият брой екзацербации в кълъстер 3 (1.84 срещу 1.28 средно за следващата година при проследяване).

**Таблица 62. Средни стойности на параметрите в отделните кълъстери (анализ с 4 кълъстера)**

Характеристика	Кълъстер 1	Кълъстер 2	Кълъстер 3	Кълъстер 4	Средна стойност за извадката
Пол, мъже	<b>88.9%</b>	47.6%	64.9%	50.0%	62.5%
Възраст, години	<b>60.3</b>	63.2	67.6	66.9	65.1
Пакетогодини тютюнопушене	23.6	27.6	<b>34.3</b>	24.0	28.7
ВМІ	25.9	25.7	26.1	27.4	26.3
Предходна диагноза бронхиална астма	16.7%	<b>23.8%</b>	5.4%	5.0%	11.5%
АСО	16.7%	<b>23.8%</b>	5.4%	10.0%	12.5%
Емфизем (фенотип)	44.4%	42.9%	<b>73.0%</b>	55.0%	57.3%
Хроничен бронхит (фенотип)	66.7%	57.1%	67.6%	60.0%	63.5%
OSA	5.6%	4.8%	8.1%	5.0%	6.3%
Артериална хипертония	50.0%	66.7%	86.5%	80%	74.0%
Дислипидемия	11.1%	28.6%	21.6%	<b>5.0%</b>	17.7%
ФЕО1 след бронходилататор, %	<b>71.4</b>	58.2	<b>45.9</b>	57.4	55.8%
Симптом: свирене в гърдите	<b>33.3%</b>	81.0%	89.2%	65.0%	71.9%
Симптом: експекторация	<b>44.4%</b>	71.4%	83.8%	75.0%	71.9%
Симптом: задух	<b>44.4%</b>	95.2%	94.6%	80%	82.3%
Умерено-тежки екзацербации в предходната	<b>0.17</b>	0.71	0.89	0.80	0.70

година					
Тежки екзацербации в предходната година	1.00	0.81	1.24	1.10	1.07
mMRC резултат	<b>0.83</b>	2.19	<b>2.78</b>	1.70	2.1
CAT резултат	<b>6.06</b>	13.9	<b>22.9</b>	12.7	15.6
Еозинофили, абсолютен брой	0.25	0.23	0.21	0.24	0.23
CRP, mg/dL	<b>2.27</b>	3.84	5.64	4.66	4.41
Изминато разстояние на 6MWT, m	<b>527.2</b>	395.0	<b>255.7</b>	425.0	372.3
Десатурация след 6MWT, %	3.3	3.5	5.8	4.4	4.5
РАPm, mmHg	<b>17.7</b>	21.0	27.1	19.3	22.4
Наличие на пулмонална хипертония	22.2%	47.6%	<b>83.8%</b>	25.0%	52.1%
Подвижност на дясна диафрагма в покой, mm	<b>58.1</b>	48.1	42.3	49.7	48.1
Общ брой екзацербации след 1 година	0.83	1.10	<b>1.84</b>	0.85	1.28

При сравняване на резултатите с йерархичната клъстеризация, методът на *K*-средните отново е свързан с по-ниски разлики (по-лоши резултати) между групите по отношение на отделните параметри и при прогнозиране на екзацербациите в следващата година.

### Йерархична клъстеризация (Hierarchical Clustering)

За установяване на оптималния брой на клъстерите са използвани агломеративните коефициенти, което позволява обективизиране на методите за йерархична клъстеризация за идентифициране на по-хомогенни клъстерни структури. Стойности, близки до „1“, предполагат силна клъстерна структура. На Таблица 63 са показани агломеративните коефициенти на различните методи. С най-висока стойност е методът „Ward.D“.

**Таблица 63. Агломеративни коефициенти на различните методи**

Метод	Агломеративен коефициент
Complete	0.483
Average	0.364
Single	0.300
Ward.D	0.746
Mcquitty	N/A
Centroid	N/A

Подобно на начина за определяне на оптималния брой кълъстери при метода на *K*-средните, при йерархична кълъстеризация също може да се приложи сходен подход с аналогичен резултат.

Проведеният кълъстерен анализ показва наличието на възможни клинични кълъстери, свързани с различна честота на екзацербации. Разделянето на извадката на два кълъстера обособява следните групи (Таблица 64):

- в кълъстер 1 (АСО фенотип) попадат предимно пациенти с предходна диагноза бронхиална астма и АСО, с по-леко заболяване (ФЕО1) и по-малко симптоми, по-добро качество на живот (САТ и mMRC) и по-добра диафрагмална подвижност.
- в кълъстер 2 (не-АСО фенотип) попадат пациенти с авансирала ХОББ (по-нисък ФЕО1) с наличие както на емфиземен, така и на бронхитен фенотип, значими коморбидности (OSA, артериална хипертония, дислипидемия) и по-нисък физически капацитет (6MWT).

Трябва да се отбележи, че разликата в ВМІ между двата кълъстера е 0.7 (25.9 срещу 26.6).

**Таблица 64. Средни стойности на параметрите в отделните кълъстери при йерархична кълъстеризация (анализ с 2 кълъстера)**

Характеристика	Кълъстер 1	Кълъстер 2	Средна стойност за извадката
Предходна диагноза бронхиална астма	19.1%	4.1%	11.5%
АСО	21.3%	4.1%	12.5%
Емфизем (фенотип)	44.7%	69.4%	57.3%
Хроничен бронхит (фенотип)	55.3%	71.4%	63.5%
OSA	2.1%	10.2%	6.3%

Артериална хипертония	57.4%	89.8%	74.0%
Дислипидемия	6.4%	28.6%	17.7%
ФЕО1 след бронходилататор, %	66.8%	45.2%	55.8%
Симптом: свирене в гърдите	48.9%	93.9%	71.9%
Симптом: експекторация	55.3%	87.8%	71.9%
Симптом: задух	66.0%	98.0%	82.3%
Умерено-тежки екзацербации в предходната година	0.49	0.90	0.70
Тежки екзацербации в предходната година	0.96	1.18	1.07
mMRC резултат	1.32	2.78	2.1
CAT резултат	9.0	22.0	15.6
Еозинофили, абсолютен брой	0.27	0.18	0.23
CRP, mg/dL	3.73	5.08	4.41
Изминато разстояние на 6MWT, m	435.6	311.6	372.3
Десатурация след 6MWT, %	3.1	5.9	4.5
PAPm, mmHg	17.7	26.9	22.4
Наличие на пулмонална хипертония	19.1%	83.7%	52.1%
Подвижност на дясна диафрагма в покой, mm	54.1	42.3	48.1
Общ брой екзацербации след 1 година	0.77	1.78	1.28

Разделянето на извадката на четири клъстера обособява следните групи (Таблица 65):

- в клъстер 1 (неекзацербатори) попадат предимно пациенти с добро качество на живот (CAT и mMRC), по-добра белодробна функция (ФЕО1), съхранена физическа активност (6MWT) и ниска честота на екзацербации в предходната година;
- в клъстер 2 (емфиземен фенотип) попадат пациенти с авансирала ХОББ (по-нисък ФЕО1) с наличие на емфиземен фенотип, пулмонална хипертония, намалена диафрагмална подвижност и по-нисък физически капацитет (6MWT).
- в клъстер 3 (АСО фенотип) попадат предимно по-млади жени с предходна диагноза бронхиална астма и АСО, с по-леко заболяване (ФЕО1), висок еозинофилен брой,

сравнително добро качество на живот (CAT и mMRC) и по-добра диафрагмална подвижност.

- в кълъстер 4 (бронхитен фенотип) попадат пациенти с висок BMI, бронхитен фенотип, значими коморбидности (OSA, артериална хипертония, дислипидемия), влошено качество на живот (CAT и mMRC) и сравнително по-високи маркери на възпаление (CRP).

**Таблица 65. Средни стойности на параметрите в отделните кълъстери при йерархична кълъстеризация (анализ с 4 кълъстера)**

Характеристика	Кълъстер 1	Кълъстер 2	Кълъстер 3	Кълъстер 4	Средна стойност за извадката
Пол, мъже	71.1%	61.1%	<b>11.1%</b>	76.9%	62.5%
Възраст, години	63.8	67.2	<b>60.7</b>	66.1	65.1
Пакетогодини	25.0	34.8	<b>15.6</b>	31.8	28.7
BMI	25.9	24.9	25.8	<b>31.3</b>	26.3
Предходна диагноза бронхиална астма	5.3%	5.6%	<b>77.8%</b>	0%	11.5%
АСО	5.3%	5.5%	<b>88.9%</b>	0%	12.5%
Емфизем (фенотип)	47.4%	<b>80.1%</b>	33.3%	38.5%	57.3%
Хроничен бронхит (фенотип)	60.5%	66.7%	33.3%	<b>84.6%</b>	63.5%
OSA	2.6%	0%	0%	<b>38.5%</b>	6.3%
Артериална хипертония	55.3%	86.1%	66.7%	<b>100%</b>	74.0%
Дислипидемия	7.9%	19.4%	0%	<b>53.8%</b>	17.7%
ФЕО1, %, след бронходилататор	68.2	<b>41.3</b>	61.1	56.0	55.8%
Свирене в гърдите (симптом)	<b>39.5%</b>	97.2%	88.9%	84.6%	71.9%
Експекторация (симптом)	<b>52.6%</b>	83.3%	66.7%	100%	71.9%
Задух (симптом)	<b>60.5%</b>	97.2%	88.9%	100%	82.3%
Умерено-тежки	<b>0.39</b>	1.03	0.89	0.54	0.70

екзацербации в предходната година					
Тежки екзацербации в предходната година	<b>0.76</b>	1.13	1.78	1.31	1.07
mMRC	<b>1.2</b>	2.8	2.0	2.8	2.1
CAT	<b>7.9</b>	21.7	13.6	<b>22.8</b>	15.6
Еозинофилен брой	0.23	0.16	<b>0.44</b>	0.25	0.23
CRP	3.82	4.88	3.32	<b>5.62</b>	4.41
Изминато разстояние на 6MWT, m	<b>448.4</b>	<b>303.8</b>	381.7	333.5	372.3
Десатурация след 6MWT, %	3.1	7.0	2.9	2.8	4.5
РАРm, mmHg	17.7	<b>28.1</b>	17.4	23.5	22.4
Наличие на пулмонална хипертония	21.1%	<b>91.7%</b>	<b>11.1%</b>	61.5%	52.1%
Подвижност на дясна диафрагма в покой, mm	54.3	<b>39.8</b>	53.0	49.4	48.1
Общ брой екзацербации след 1 година	<b>0.34</b>	1.86	<b>2.56</b>	1.54	1.28

Наблюдават се значими разлики в честотата на екзацербациите между кълстерите, при разделяне на групата на два кълстера (0.77 срещу 1.78 обостряния) и четири кълстера (максимална честота от 2.56 в кълстер 3, минимална честота от 0.34 екзацербации в кълстер 1, със средна честота от 1.28 екзацербации в цялата извадка).

## 5. Дискусия

Настоящият научен труд демонстрира възможностите от приложението на изкуствен интелект и по-специално - машинно обучение, при пациенти с ХОББ върху прогнозирането на клинично значими крайни точки и анализира резултатите спрямо настоящите ръководства и препоръки, където това е възможно.

В първата част на докладваните резултати е проведена дескриптивна статистика с описание на променливите, разделени по групи (анамнестични показатели; честота на предходни екзацербации; спирометрични показатели; качество на живот и симптоми; хематологични показатели; ехокардиографски показатели; ехографски показатели; ЕКГ показатели; оценка на физическия капацитет), и последващ еднокомпонентен анализ (изследване на корелации между отделните променливи, определяне на оптимална гранична стойност и прогностична стойност на отделна променлива при прогнозиране на честотата на екзацербациите и композитна крайна точка).

Демонстрирана е възможността за установяване на оптимална гранична точка на отделна характеристика, която е свързана с най-висока прогностична стойност. В случаите, където такава е налична, е оценена нейната точност. В този контекст настоящото проучване потвърждава оптимална гранична точка на характеристиката PAr<sub>m</sub> от 20 mmHg за установяване на повишен риск от екзацербации. Настоящите препоръки за диагностициране на пулмонална хипертония препоръчват същата стойност (234). В случаите, където няма консенсусна крайна точка (например диафрагмална подвижност, някои ехокардиографски параметри), е демонстриран начинът на установяването на такава, заедно с методите за оценка на нейната информативност.

Настоящото проучване установява прогностичната стойност на редица параметри. Като новаторски аспект то за първи път характеризира влиянието на **диафрагмалната дисфункция** при протичането на ХОББ. Нещо повече, това е първото проучване, което изследва промяната на диафрагмалната подвижност при усилие и връзката му с редица параметри на заболяването. Въпреки че диафрагмалната подвижност в покой е свързана с много аспекти на заболяването, разликата в диафрагмалната подвижност след усилие допълнително характеризира протичането, особено по отношение на системното възпаление (CRP).

Диафрагмалната дисфункция е честа при пациенти с ХОББ. Тя влошава задуха, физическия капацитет и намалява качеството на живот. Пациентите с ХОББ имат намален физически капацитет и като последица от нарушената белодробна функция. Диафрагмалната подвижност в покой показва значима връзка с ФЕО1, фенотипа на честия екзацербатор, честотата на умерено-тежки екзацербации, качеството на живот и наличието на пулмонална хипертония. Аналогично, спадът на диафрагмалната подвижност след усилие показва значима връзка с ФЕО1, фенотипа на честия екзацербатор, честотата на тежките екзацербации, качеството на живот и наличието на пулмонална хипертония. В този смисъл пулмоналната рехабилитация може да подобри състоянието на пациентите с ХОББ, при които е налице диафрагмална дисфункция. Неинвазивната оценка на диафрагмалната функция чрез ехографско измерване на движението на диафрагмата в покой и след усилие може да допринесе за оценката на тежестта на заболяването.

Диафрагмата при здрав човек има средна екскурзия (амплитуда на движение) 63 mm, докато при пациенти с ХОББ – 23 mm, като това намаление е правопрпорционално на ФЕО1 (457). Друго проучване също показва, че пациентите с ХОББ имат по-малка амплитуда на диафрагмалните екскурзии в сравнение с контролна група (36 mm спрямо 46 mm), като е налице зависимост с изминатото разстояние при 6MWT (458). В допълнение, свръхраздуването води до изместване от оптималната позиция, което намалява генерираното налягане при работа. Не всички проучвания обаче намират разлика в диафрагмалните екскурзии при пациенти с ХОББ спрямо здрави контроли (459).

Връзката между диафрагмалната подвижност, спада й след усилие и ФЕО1 е очаквана и описана с оглед динамичното свръхраздуване (457,458). Останалите прояви вследствие на диафрагмалната дисфункция обаче заслужават внимание. Честотата на екзацербациите е по-честа при влошена диафрагмална функция. Пулмоналната хипертония е по-честа при намалена диафрагмална подвижност в покой, а спад на диафрагмалната подвижност с  $\geq 1$  mm след усилие е сигнификантен маркер за наличието й. За отбелязване е и фактът, че диафрагмалната подвижност корелира с PAr<sub>m</sub> ( $r = -0.6$ ) в най-висока степен от всички изследвани параметри, дори в сравнение с ФЕО1 ( $r = -0.46$ ). Лабораторните маркери не показват връзка със спада в диафрагмалната

подвижност, но пациентите със значителен спад на диафрагмалната подвижност след усилие ( $\geq 17$  mm) имат значимо по-високи стойности на CRP, предполагащо по-тежко възпаление.

Трябва да се има предвид, че „оптималната гранична точка“ е изчислена на базата на извадката в проучването. По този начин изчислените стойности могат да не са оптимални в популацията или в друга извадка. Това налага последващи проучвания да валидират представените резултати.

Към настоящия момент GOLD препоръчва броят на екзацербациите в предходната година като предиктор за честотата на екзацербациите в следващата година. Нови данни показват обаче, че анамнезата за екзацербации може да е компрометирана като рисков индикатор за ХОББ (460). Клиничните ръководства препоръчват анамнезата за екзацербации като основен фактор за избора на терапия и за прогнозиране на риска от екзацербации на ХОББ, но анализ на данни от три различни клинични проучвания установява, че анамнезата за екзацербации сама по себе си не е най-точният инструмент за прогнозиране на риска. Сегашното определение на фенотипа на честия екзацербатор се основава на схващането, че най-добрият предиктор за бъдещи екзацербации са предходните екзацербации. Анамнезата за екзацербации вероятно е най-добрият индивидуален предиктор за бъдещи екзацербации, но не е достатъчно добра, за да служи като основа за надеждно определение на фенотипа, което да дава информация за избора на лечение. За да се установи по-стабилен фенотип на екзацербациите, е необходимо да се премине от настоящия еднофакторен подход към по-широка класификация на риска, основана на множество характеристики на пациента и заболяването. Нарастващите данни от кохорти на ХОББ, съчетано с употребата на статистически методи, би трябвало да постигне тази цел, което от своя страна ще доведе до по-стабилни и последователни стратегии за клинично поведение при пациентите.

Проучването IMPACT (460) анализира три изследвания, в които са включени 4107 пациенти с различни нива на риск за умерени или тежки екзацербации:

- плацебо рамото на проучването Study to Understand Mortality and Morbidity in COPD (SUMMIT; N = 2421);
- проучването Long-term Oxygen Treatment Trial (LOTT; N = 595);

- плацебо рамото на проучването Towards a Revolution in COPD Health (TORCH; N = 1091).

Рискът от екзацербация на заболяването е съответно нисък, среден и висок в трите изпитвания. Стратификацията на риска е осъществена чрез сравняване на ефективността на три алгоритъма:

- анамнеза за екзацербации;
- моделът, разработен от д-р Бертенс и колегите му в Нидерландия през 2013 година (461);
- последната версия на инструмента за прогнозиране на екзацербации на ХОББ, известен като АСCEPT (462).

За оценка на прогностичната ефективност на алгоритмите в проучването е използвана площта под кривата (AUC). АСCEPT превъзхожда анамнезата за екзацербации и алгоритъма на Бертенс във всички групи в LOTT (среден риск) и TORCH (висок риск), като и в двата случая резултатите са статистически значими. В SUMMIT (нисък риск) Бертенс алгоритъмът и АСCEPT статистически значимо превъзхождат анамнезата за екзацербации. AUC при употреба само на анамнезата за екзацербации при прогнозиране на бъдещи обостряния в SUMMIT, LOTT и TORCH е съответно 0.59 (95% CI 0.57 – 0.61), 0.63 (95% CI 0.59 – 0.67) и 0.65 (95% CI 0.63 – 0.68). Бертенс алгоритъмът има по-висока AUC в сравнение със самостоятелна анамнеза на обострянията в SUMMIT (увеличение с 0.10,  $P < 0.001$ ) и TORCH (увеличение с 0.05,  $P < 0.001$ ), но не и в LOTT (увеличение с 0.01,  $P = 0.84$ ). АСCEPT алгоритъмът има по-висока AUC в сравнение само с анамнезата за екзацербации във всички изследвани групи, съответно с 0.08 ( $P < 0.001$ ), 0.07 ( $P = 0.001$ ) и 0.10 ( $P < 0.001$ ). В сравнение с алгоритъма на Бертенс, АСCEPT има по-висока AUC с 0.06 ( $P = 0.001$ ) в LOTT и 0.05 ( $P < 0.001$ ) в TORCH, докато AUC не се различават в SUMMIT (промяна от -0.02,  $P = 0.16$ ).

Между 29 и 47% от пациентите с ХОББ изпитват поне една екзацербация годишно (463). Различията в клиничните характеристики и системните прояви превръщат ХОББ в хетерогенно заболяване, което се свързва с различни рискове от екзацербации, хоспитализации и смъртност. В тази връзка **фенотипизирането** на ХОББ може да допринесе за по-точна оценка на този риск. Фенотипите на ХОББ зависят от характеристиките на оценяваната популация като тютюнопушене, респираторни симптоми, съпътстващи заболявания, спирометрия, физически капацитет и възпалителни биомаркери (464–466).

През годините са извършвани много промени на съществуващи алгоритми и предложения за различни класификации. Най-значимата от тях е въвеждането на класификацията A-B-C-D на GOLD през 2011 година. През 2023 г. GOLD обединява групите C и D в една група, наречена "E", за да подчертае клиничното значение на екзацербациите (35).

Проведеният клъстерен анализ в настоящия научен труд показва по-добри резултати на йерархичната кластеризация при прогнозиране на риска от екзацербации. Значими разлики в честотата на екзацербациите между клъстерите се наблюдават при разделяне на групата на два клъстера (0.77 срещу 1.78 обостряния) и четири клъстера (максимална честота от 2.56 в клъстер 3, минимална честота от 0.34 екзацербации в клъстер 1, със средна честота в цялата извадка от 1.28 екзацербации).

Няколко проучвания също използват клъстерен анализ за изследване на кохорти от пациенти с ХОББ с цел идентифициране на клинични фенотипи (467–475). *Burgel et al.* идентифицират пациенти с ХОББ с тежко ограничение на въздушния поток в ранна възраст в две различни кохорти (468,469). Тези пациенти се характеризират с кахексия (468,469), емфизем и висока честота на екзацербации на ХОББ (469), мускулна слабост и остеопороза (469), но с много ниска честота на сърдечносъдови съпътстващи заболявания (468,469). И в двете проучвания този фенотип на заболяването е свързан с много висок риск от смърт в сравнително млада възраст (468,469), което налага специфична терапевтична намеса при тази група. Такава група не е обособена в клъстерния анализ на това проучване и се припокрива в най-голяма степен с клъстер 2 (висока честота на емфизем, нарушена белодробна функция) и клъстер 3 (ниска сърдечносъдова коморбидност, висока честота на екзацербации).

АСО фенотипът е добре позната клинична група, като тези пациенти са по-млади, с по-малко пакетогодини тютюнопушене и по-често са жени в сравнение с другите фенотипи (476). Освен това пациентите с АСО имат относително запазено общо състояние в сравнение с други фенотипи, както по отношение на BODE индекса, така и по отношение на степента на обструкция. Това съответства в най-голяма степен на клъстер 3 в представения анализ. За отбелязване е, че това е и клъстерът с най-висока честота на екзацербации, въпреки относително запазената белодробна

функция, ниското разпространение на тютюнопушенето и по-младата възраст.

В някои проучвания са идентифицирани групи пациенти с ХОББ с метаболитни и сърдечносъдови съпътстващи заболявания. *Garcia-Aymerich et al.* идентифицират група от пациенти със „системна“ ХОББ (467). Тези участници се характеризират с висок индекс на телесна маса и много висока честота на диабет, застойна сърдечна недостатъчност и исхемична болест на сърцето. Интересно е, че те имат по-високи нива на задух и влошен здравен статус спрямо участниците със сравнимо ограничение на въздушния поток, но с по-малко сърдечносъдови и метаболитни съпътстващи заболявания (467). Важно е да се отбележи, че тези пациенти са имали висок риск от хоспитализация поради сърдечносъдови събития и значителен риск от хоспитализация поради ХОББ и обща смъртност, въпреки че са имали умерено ограничение на въздушния поток (467). Това описание най-много наподобява клъстер 4 в представения анализ.

Този анализ демонстрира значимостта на АСО фенотипа при оценката на риска от екзацербации. Интересно е да се отбележи, че анализът с два клъстера не разделя изследваната група на типичните фенотипи на ХОББ - емфиземен и бронхитен, а предпочита разделянето на групите на такива с висока и ниска честота на АСО. Емфиземната и бронхитната ХОББ са разграничени като клъстер 2 и клъстер 4 само при четириклъстерния анализ. През 2014 г. GOLD припознава АСО фенотипа, но през 2020 г. вече не използва този термин, въпреки практическата му стойност (477), както показва и този анализ.

В последните години разбирането на факторите, които допринасят за хетерогенността на заболяването, клиничните прояви и отговора към терапията, е значително подобро. Съответно, диагностичната оценка и клиничното поведение при ХОББ се превръщат в по-индивидуализирани. Въпреки тези постижения, настоящото клинично поведение при ХОББ и прогнозирането на риска, както се препоръчва от Глобалната инициатива за хронична обструктивна белодробна болест (GOLD), все още е далеч от това да бъде наистина „персонализирано“.

Интересът към прилагането на ИИ и МО нараства във всички области на медицината, което може да улесни персонализирането на подхода (429). МО може да осигури значителен напредък в прилагането на персонализираната

медицина при ХОББ, като изчислява точен риск за конкретни крайни точки (например екзацербация, смърт и т.н.) при отделните пациенти (въз основа на параметри като белодробна функция, клинични и демографски променливи, предходни екзацербации, лабораторни биомаркери като еозинофили и др.).

През 2012 г. *European Respiratory Journal* публикува една от първите ключови статии, в която се обсъжда необходимостта от персонализирана медицина при екзацербациите на ХОББ (478). Разпознаването на този подход е развито, за да се формират единици, към които може да се насочи вниманието (т.е. лечими черти, *treatable traits*) (479,480). Въпреки това, след тази последна “state-of-art” актуализация, е постигнат много малък напредък в съответната насока (481). От друга страна МО показва обещаващи резултати в това направление и би могло да помогне за идентифициране на респондерите и нереспондерите (или дори суперреспондерите) на определена интервенция (482). **Настоящият труд обобщава възможностите за персонализирана медицина при ХОББ и посочва възможностите за прилагане на МО.**

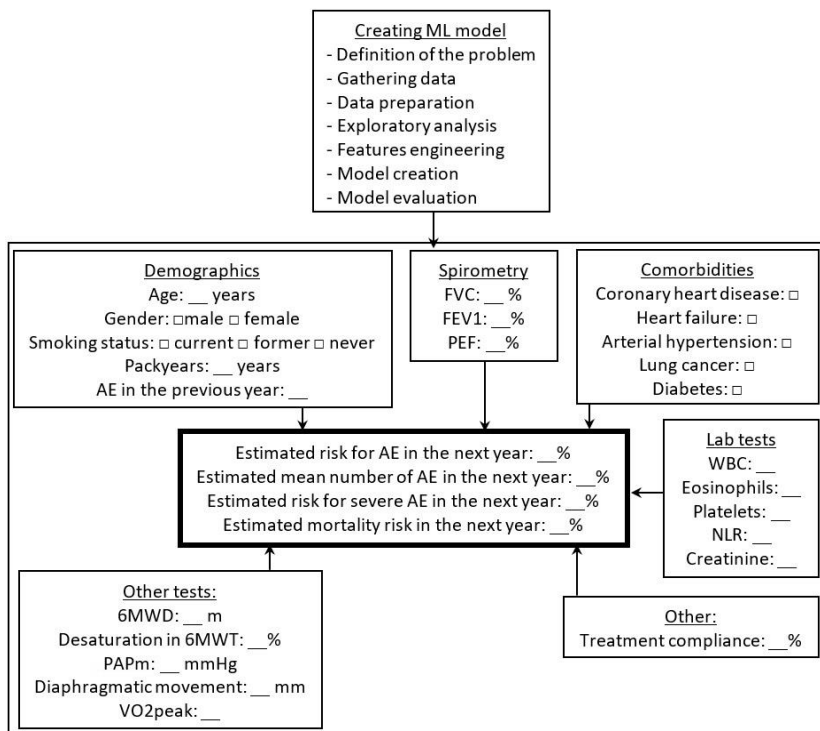
Извършването на информирана клинична оценка въз основа на съществуващите данни е основен принцип на медицината, базирана на доказателства. През последните години МО бързо се развива в медицината, като първите стъпки са направени и в респираторната медицина (429,483). От създаването на регресионния модел през 1805 г. обработката на данни непрекъснато се развива - от прости методи като дървета на решенията (*decision trees*) (1966 г.) до *Random Forest* (2001 г.). Понастоящем *boost* моделите (например *XGBoost* през 2014 г. и *CatBoost* през 2017 г.) са най-съвременните (*state-of-art*) и са обичайно един от първите избори. Най-новата тенденция в МО е прогнозиране чрез комбинация от няколко модела (т. нар. ансамблови модели). Ансамбловите методи съчетават статистика и МО, като използват набор от алгоритми за обучение, за да постигнат по-добра точност на прогнозиране, отколкото всеки от алгоритмите поотделно (484).

Независимо от модела, **основните етапи за създаване на МО алгоритъм са:**

- Дефиниране на проблема;
- Събиране на данни;
- Подготовка на данните;
- Проучвателен (експлораторен) анализ;

- Конструирание на променливите (features engineering);
- Създаване на модел;
- Оценка на модела.

Въпреки че създаването на модел изисква значителни познания в областта и умения за работа с данни, понастоящем се появяват автоматизирани рамки за МО (AutoML) (например Google Vertex AI, Microsoft Azure, Data Robot AI), които създават възможност за изграждане, обучение и внедряване на авангардни МО модели, без да е необходимо писане на код. Освен това могат да се разработят леснодостъпни надеждни и динамични интерфейси, които да адаптират оформлението и функционалността на разработените модели. Те могат да бъдат представени като празни полета за всеки параметър, които да се попълват от здравния работник (Фигура 44), и може лесно да бъде постигнато на различни платформи (например компютър, смартфон, таблет).



**Фигура 44. Пример за МО модел за оценка на едногодишен риск**

Съкращения: AE - екзацербация; FVC - форсиран витален капацитет; FEV1 - форсиран експираторен обем за първата секунда; PEF - върхов експираторен дебит; WBC - левкоцити; NLR - съотношение неутрофили-лимфоцити; 6MWD - изминато разстояние при шестминутен тест с ходене; 6MWT - шестминутен тест с ходене; PAPm - средно налягане в белодробната артерия; VO2peak - пиково поглъщане на кислород

В литературния обзор е проведен цялостен анализ на предиктивните фактори (демографски фактори, честота на предходни екзацербации, спирометрични показатели, качество на живот и симптоми, лабораторни показатели, ехокардиографски, ехографски и ЕКГ показатели и оценка на физическия капацитет). Алгоритъмът на МО може да използва всички традиционни рискови фактори, но също така да включва и други, които могат да бъдат специфични за конкретната обстановка (например рискът от екзацербация може да бъде свързан със замърсяването на околната среда, но той варира значително в различните държави или в различните региони на една държава). Стандартен инструмент за прогнозиране, като например *DECAF score*, не може да обхване тези нюанси, докато МО може да оцени всички локални фактори и да генерира персонализирана прогноза на риска, като вземе предвид например околната среда на пациента, анамнезата, жилищните условия, сърдечна честота в покой, статуса на ваксинация срещу грип/пневмококи/COVID-19 и т.н.

Възможностите на МО, свързани с отстраняването на проблеми в набора от данни (например липсващи данни) също са значителни. Обработката на липсващи данни е от решаващо значение при подготовката на набори от данни за модели за машинно обучение, тъй като липсващите данни могат значително да повлияят на ефективността на тези модели. Изборът на подходящ метод зависи от естеството и количеството на липсващите данни, очакваното въздействие на липсващите данни върху модела и вида на използвания модел. Често е полезно да се изпробват няколко метода и да се сравнят техните резултати чрез техники за валидиране на модела. Някои от по-често срещаните методи за въвеждане на липсващи данни включват:

- Изтриване - изтриване на целия запис/ред (премахване на цели наблюдения, в които липсва каквато и да е стойност) или сдвоено изтриване (премахване на конкретни липсващи точки от данни, но запазване на останалите данни в наблюдението).

- Импутация – замяна на липсващите стойности със средната стойност, медианата или модата на наблюдаваните стойности в променливата. В някои случаи може да се използва импутация на случайна стойност (замяна на липсващите стойности с произволно избрани наблюдавани стойности), импутация на най-близките съседни стойности (замяна на липсващите стойности със средната стойност или медианата на k-най-близките наблюдавани стойности), интерполация и екстраполация (замяна на липсващите стойности чрез интерполация или екстраполация от наблюдаваните стойности).

- Импутация, базирана на модел - регресионна импутация (замяна на липсващите стойности с помощта на регресионен модел), стохастична регресионна импутация (подобна на регресионната импутация, но с добавяне на случайна остатъчна стойност към всяка импутация), множествена импутация (създаване на множество набори от данни с различни импутирани стойности, анализирани поотделно и след това осредняване на резултатите).

- Deep learning техники - автокодери и генератори на данни (невронни мрежи, които генерират нови данни, подобни на входните данни).

- Индикаторна променлива - добавяне на нова променлива, указваща дали липсват данни за дадено наблюдение, след което се присвояват липсващите данни (често с „нула“).

- Хибридни методи - комбиниране на различни методи за импутация.

В настоящия научен труд при липсващи данни е използвана замяна със средната стойност и медианата в зависимост от разпределението.

МО може да бъде следващата стъпка на персонализираната медицина при ХОББ, като дава точния риск (риск от обостряне, смърт и т.н.) на всеки пациент (въз основа на неговите параметри като белодробна функция, клинични данни, демографски данни, предходни екзацербации и т.н.) и по този начин предоставя прогноза/риск за конкретния пациент въз основа на индивидуалните му характеристики (**индивидуален подход**). Това е стъпка напред в сравнение с обичайно използваното разпределяне на пациентите в конкретна група, за която са налични данни за прогноза/риск (групов подход).

### 5.1. Избор на най-добър модел за прогнозиране наличие на екзацербация

Сравнението на точността на моделите за прогнозиране наличие на екзацербация показва най-добри резултати за KNN (82.8%), Random Forest (82.8%) и Naïve Bayes (82.8%) моделите (Таблица 66).

**Таблица 66. Сравнение на моделите за прогнозиране наличие на екзацербация**

Модел	Нулев модел	Логистична регресия	KNN	Дърво на решението	Random Forest
Точност	65.5% (19/29)	65.5% (19/29)	<b>82.8% (24/29)</b>	79.3% (23/29)	<b>82.8% (24/29)</b>
Модел	Naïve Bayes	Невронна мрежа	SVM	XGBoost	CatBoost
Точност	<b>82.8% (24/29)</b>	65.5% (19/29)	79.3% (23/29)	79.3% (23/29)	79.3% (23/29)

Ансамбловият модел за прогнозиране наличие на екзацербация също класифицира правилно 82.8% (24/29) от наблюденията.

### 5.2. Избор на най-добър модел за прогнозиране на риск по GOLD

Сравнението на точността на моделите за прогнозиране на риск по GOLD показва най-добри резултати за логистична регресия (75.9%), следвани от CatBoost (72.4%) и дърво на решението (69.0%) и Random Forest (69.0%) моделите (Таблица 67).

**Таблица 67. Сравнение на моделите за прогнозиране на риск по GOLD**

Модел	Нулев (GOLD) модел	Логистична регресия	KNN	Дърво на решението	Random Forest
Точност	58.6% (17/29)	<b>75.9% (22/29)</b>	62.1% (18/29)	69.0% (20/29)	69.0% (20/29)
Модел	Naïve Bayes	Невронна мрежа	SVM	XGBoost	CatBoost
Точност	55.2%	62.1% (18/29)	55.2%	65.5%	72.4%

(16/29)	(16/29)	(19/29)	(21/29)
---------	---------	---------	---------

Ансамбловият модел за прогнозиране на риск по GOLD класифицира правилно 72.4% (21/29) от наблюденията.

### 5.3. Избор на най-добър модел за честота на екзацербации (непрекъсната променлива)

Сравнението на RMSE на моделите за прогнозиране на честотата на екзацербациите като непрекъсната променлива показва най-добри резултати за логистична регресия (0.970), следвани от Random Forest (0.989) и CatBoost (1.010) (Таблица 68).

**Таблица 68. Сравнение на моделите за прогнозиране на честотата на екзацербациите (непрекъсната променлива)**

Модел	Нуле в модел	Линей на регресия	Дърво на решение	Random Forest	Неврон на мрежа	XGBoost	CatBoost
RMS E	1.100	<b>0.970</b>	1.329	0.989	1.028	1.174	1.010

Ансамбловият модел за прогнозиране на броя на екзацербациите (категорийна променлива) показва RMSE от 0.959.

### 5.4. Избор на най-добър модел за честота на екзацербации (категорийна променлива)

Сравнението на точността на моделите за прогнозиране на честотата на екзацербациите като категорийна променлива показва най-добри резултати за CatBoost модела (48.3%), следвани от Random Forest (44.8%) и XGBoost (34.5%) (Таблица 69).

**Таблица 69. Сравнение на моделите за прогнозиране на честотата на екзацербациите (категорийна променлива)**

Модел	Нуле в модел	Логистична регресия	Дърво на решение	Random Forest	Неврон на мрежа	XGBoost	CatBoost
Точно	34.5	31.0%	31.0%	44.8%	34.5%	37.9%	<b>48.3%</b>

ст	% (10/29)	(9/29)	(9/29)	(13/29)	(10/29)	(11/29)	<b>(14/29)</b>
----	--------------	--------	--------	---------	---------	---------	----------------

Ансамбловият модел за прогнозиране на броя на екзацербациите като категорична променлива класифицира правилно 41.4% (12/29) от наблюденията.

## 6. Заключение

Извършването на информирана клинична преценка въз основа на съществуващите данни е в основата на медицината, основана на доказателства. Възможността за определяне на максимално информативна гранична стойност в зависимост от зададената цел (минимизиране/максимизиране) позволява както оценка на настоящите препоръки за съответния показател, така и предложение на по-добри алтернативи.

Моделите, използващи МО за прогнозиране на риск от екзацербация и смърт при ХОББ превъзхождат настоящите ръководства и в почти всички случаи традиционните регресионни модели. Възможностите за създаване на модел за прогнозиране на индивидуален риск в зависимост от наличността на показателите и/или извадка при повишаване на точността показва предимствата на методиката.

Фенотипният анализ посредством K-means клъстеризация и йерархична клъстеризация показва показателите, които според МО имат най-голяма важност при обособяване на групи (фенотипи).

Алгоритмите за МО може да използват редица традиционни рискови фактори, но също така да включват и други, които могат да бъдат специфични за конкретното място (например рискът от екзацербация може да бъде свързан със замърсяването на околната среда, но може да варира значително в различните държави или в различните региони на една държава).

МО може да се превърне в следващата стъпка на персонализираната медицина при ХОББ, като дава точния риск (риск от обостряне, смърт и т.н.) на всеки пациент (въз основа на параметри като белодробна функция, клинични данни, демографски данни, предходни екзацербации и т.н.) и по този начин предоставя прогноза/риск за конкретния пациент въз основа на индивидуалните му характеристики (индивидуален подход). Това е стъпка напред в сравнение с обичайно използваното разпределяне на пациентите в конкретна група, за която са налични данни за прогноза/риск (групов подход). Трябва изрично да се подчертае, че всички тези нови разработки в областта на ИИ и МО биха могли да помогнат за вземането на клинични решения, но не могат да заменят напълно лекаря. По този начин крайната отговорност за решенията остава в ръцете на клинициста.

## **Приноси**

### Приноси с научно-теоретичен характер

1. Проведен е детайлен литературен обзор върху рисковите фактори, свързани с влошаване на прогнозата при ХОББ (екзацербации, смърт);
2. Направена е характеристика с подробно описание на широк набор от показатели (анамнестични данни/демографски показатели, спирометрични показатели, качество на живот, хематологични, ехокардиографски, ехографски, електрокардиографски показатели и тестове с натоварване) на популация от пациенти с ХОББ;
3. Създадена е концептуална рамка за оценка на риска при различни крайни точки. Предложеният начин на работа не се ограничава само до ХОББ като заболяване и екзацербациите като крайна точка, а може да бъде използван при различни заболявания за прогноза на всякакви показатели, които могат да придобият числов израз.

### Приноси с научно-приложен характер

1. Проведен е анализ и е създаден модел с машинно обучение за прогнозиране на риска от екзацербация или смърт, използвайки единични параметри при пациенти с ХОББ;
2. Проведен е анализ и са създадени модели с машинно обучение за прогнозиране на риска от екзацербация или смърт, използвайки съвкупност от параметри при пациенти с ХОББ;
3. Проведен е анализ на резултатите от създадените модели един спрямо друг и спрямо настоящите препоръки, където е възможно;
4. Проведен е фенотипен анализ посредством няколко различни метода. Различните клъстери, идентифицирани в тази популация от пациенти с ХОББ, са свързани с различни клинични прояви, съпътстващи заболявания и честота на екзацербациите.

## ПУБЛИКАЦИИ И УЧАСТИЯ В НАУЧНИ ФОРУМИ ПО ТЕМАТА НА ДИСЕРТАЦИЯТА

### Публикации

1. Nuñez A, Marras V, Harlander M, **Меков Е**, Esquinas C, Turel M, et al. Association Between Routine Blood Biomarkers and Clinical Phenotypes and Exacerbations in Chronic Obstructive Pulmonary Disease. *Int J Chron Obstruct Pulmon Dis* 2020;15:681-90. [IF=3.274]
2. **Меков Е**, Miravitlles M, Petkov R. Artificial intelligence and machine learning in respiratory medicine. *Expert Rev Respir Med* 2020;14(6):559-64. [IF=2.725]
3. Nuñez A, Marras V, Harlander M, **Меков Е**, Turel M, Petkov R, et al. Clinical and spirometric variables are better predictors of COPD exacerbations than routine blood biomarkers. *Respiratory Medicine* 2020;171:106091. [IF = 3.095]
4. **Меков Е**, Yanev N, Kurtelova N, Mihalova T, Tsakova A, Yamakova Y, Miravitlles M, Petkov R. Diaphragmatic Movement at Rest and After Exertion: A Non-Invasive and Easy to Obtain Prognostic Marker in COPD. *Int J Chron Obstruct Pulmon Dis*. 2022;17:1041-1050. [IF = 3.355]
5. **Меков Е**, Ilieva V. Machine learning in lung transplantation: Where are we? *Presse Med* 2022;51(4):104140 [IF = 2.7].
6. **Меков Е**, Miravitlles M, Topalovic M, Singanayagam A, Petkov R. Stepping Up the Personalized Approach in COPD with Machine Learning. *Current Respiratory Medicine Reviews* 2023;19(3):165-9. [IF = 0.2].
7. **Меков Е**, Yanev N, Petkov R, Kostadinov D. Neutrophil to lymphocyte ratio in COPD. *Thoracic medicine* 2018;10(3):8-15.
8. Petkov R, **Меков Е**, Yanev N, Kostadinov D. Diaphragm dysfunction in COPD. *Thoracic medicine* 2018;10(3):22-31.
9. **Меков Е**, Yanev N, Petkov R, Kostadinov D. Platelet count as a biomarker in COPD. *Thoracic medicine* 2019;11(3-4):34-41.
10. **Меков Е**, Ямакова Й, Петков Р. Ролята на телемедицината при ХОББ: мобилни приложения и интернет. *Наука Пулмология* 2022;3:18-22.
11. **Меков Е**, Петков Р. Как „Титаник“ продължава да променя бъдещето – машинно обучение в респираторната практика. *Варненски медицински форум* 2022;11(2):198-202.

12. **Меков Е**, Yanev N, Kurtelova N, Mihalova T, Yamakova Y, Petkov R. Echocardiographic changes of the left ventricle and left atrium in COPD. *Acta Medica Bulgarica* 2022;49(4):12-7.
13. **Меков Е**, Янев Н, Куртелова Н, Михалова Т, Ямакова Й, Петков Р. Колко активни са пациентите с ХОББ? *MedInfo* 2022;22(12):14-6.
14. **Меков Е**, Боянов М. Възможности на изкуствения интелект и машинното обучение като помощен инструмент при захарен диабет. *Наука ендокринология* 2022;4:126-30.
15. **Меков Е**, Янев Н, Куртелова Н, Михалова Т, Ямакова Й, Петков Р. ЕКГ отклонения при пациенти с ХОББ: срезови данни и прогностична стойност. *MedInfo* 2023;11:10-3.
16. **Меков Е**, Янев Н, Петков Р, Ямакова Й, Костадинов Д. Еозинофили при ХОББ. *Торакална медицина* 2024;13(1):32-41.

#### **Участия в международни форуми, резюмета, публикувани в списания с импакт фактор**

1. **Меков Е**, Yanev N, Kurtelova N, Tsakova A, Petkov R. (2020). Diaphragmatic movement in COPD patients in rest and after physical activity *ERJ* 2020;56(Suppl. 64):2205.
2. Nuñez A, Marras V, Esquinas C, Harlander M, Turel M, Lestan D, **Меков Е**, Petkov R, Pirina P, Negri S, Barrecheguren M, Miravittles M. Routine blood test biomarkers and their association with exacerbations in COPD. *ERJ* 2019;54(Suppl. 63):PA2588.