ВАРНЕНСКИ СВОБОДЕН УНИВЕРСИТЕТ „ЧЕРНОРИЗЕЦ ХРАБЪР“
ФАКУЛТЕТ “СОЦИАЛНИ, СТОПАНСКИ И КОМПЮТЪРНИ НАУКИ”
КАТЕДРА “КОМПЮТЪРНИ НАУКИ”

НЕХМЕХ ФАХРИ РМЕЙТИ

ДЪЛБОКО УЧЕНЕ ПРИ РЕШЕНИЯ ЗА ИНТЕЛИГЕНТНИ ДОМОВЕ ЗА ХОРА С ПРОБЛЕМИ

АВТОРЕФЕРАТ

на дисертационен труд
за придобиване на образователна и научна степен “доктор”,
професионално направление 4.6. Информатика и компютърни науки,
докторска програма
“Информационни системи и технологии, информатика и компютърни науки”

Научни ръководители:
Проф. д-р Росица Кузманова – Маринова
доц. д-р арх. Цвета Жекова

Варна, 2026 г.

ВАРНЕНСКИ СВОБОДЕН УНИВЕРСИТЕТ „ЧЕРНОРИЗЕЦ ХРАБЪР“
ФАКУЛТЕТ “СОЦИАЛНИ, СТОПАНСКИ И КОМПЮТЪРНИ НАУКИ”
КАТЕДРА “КОМПЮТЪРНИ НАУКИ”

НЕХМЕХ ФАХРИ РМЕЙТИ

ДЪЛБОКО УЧЕНЕ ПРИ РЕШЕНИЯ ЗА ИНТЕЛИГЕНТНИ ДОМОВЕ ЗА ХОРА С ПРОБЛЕМИ

АВТОРЕФЕРАТ

на дисертационен труд
за придобиване на образователна и научна степен “доктор”,
професионално направление 4.6. Информатика и компютърни науки,
докторска програма
“Информационни системи и технологии, информатика и компютърни науки”

Научни ръководители:
Проф. д-р Росица Кузманова – Маринова
доц. д-р арх. Цвета Жекова

Рецензенти:
проф. д.н. Борислав Стоянов
проф. д.н. Даниела Борисова

Варна, 2026 г.

Дисертационният труд е обсъден и насочен за защита пред научното жури на катедра „Компютърни науки“ към факултета “Социални, стопански и компютърни науки” на Варненския свободен университет „Черноризец Храбър“, гр. Варна.

Дисертацията е с обем от 113 страници и се състои от увод, четири глави, заключение, две приложения и използвана литература. Основният текст съдържа 14 фигури и 8 таблици. Списъкът с използвана литература се състои от 50 заглавия на английски език, включително статии, монографии и части от колективни трудове, както и от електронни източници.

Авторът на дисертационния труд е докторант на самостоятелна подготовка в катедра „Компютърни науки“ към факултета “Социални, стопански и компютърни науки” на Варненския свободен университет „Черноризец Храбър“.

Защитата на дисертационния труд пред научното жури ще се проведе на 4 февруари 2026 г. от 14,30 ч. в заседателната зала на Варненския свободен университет „Черноризец Храбър“. Материалите за защитата са на разположение на сайта на университета www.vfu.bg, раздел„Докторанти“.

I. Обща характеристика на дисертационния труд

Способността на машините да разпознават и реагират на човешки емоции в естествена комуникация става все по-значима за ефективното взаимодействие човек–компютър. Въпреки значителния напредък в областта на изкуствения интелект (AI), точното идентифициране на емоции, съдържащи се в говоримия език, остава предизвикателство поради сложността на човешките изрази, контекстните зависимости и вариациите в начина на говорене. Настоящото изследване предлага двуетапен метод, който комбинира преобразуване на реч в текст (Speech-to-Text, STT) и разпознаване на емоции върху транскрибирания текст. За класифицирането на емоции като тъга, радост, любов, гняв, страх и изненада, са приложени както класически алгоритми за машинно обучение, така и модели на дълбоко обучение (LSTM). Моделите са обучени и оценени върху разнообразни набори от данни, като се използва тестване с кръстосан набор от данни, за да се потвърди тяхната устойчивост и обобщаемост. Изследването анализира и практическата приложимост на метода в контекста на интелигентни домове за възрастни хора или хора с увреждания, където автоматичното разпознаване на емоционално състояние може да осигури навременна подкрепа и да предотврати рискови ситуации. Резултатите са принос към развиващата се област на емоционалните изчислителни системи и демонстрират как дълбокото обучение и обработката на естествения език могат да бъдат използвани за създаване на AI системи, които са по-ориентирани към човека и съпричастни. Разпознаването на емоции чрез текст, извлечен от реч, предлага по-ниска изчислителна сложност и висока ефективност в реални среди — предимство, което е ключово при системи с ограничени ресурси, внедрени в интелигентни домове.

II. Обем и структура на дисертационния труд

Дисертационният труд е с обем от 113 страници. Състои се от увод, изложение в 4 глави, заключение, списък на използваните източници и приложения. Основният текст включва 14 фигури и 8 таблици. Използваната библиография съдържа 50 източника на английски език. Във връзка с темата на дисертационния труд са направени 6 публикации.

Съдържание:

- Цел и задачи
- Изследователски въпроси
- Изследователски хипотези

Глава 2. Преглед на литературата

- Общ преглед на анализа на емоции в умен дом
- Машинно обучение и дълбоко обучение за разпознаване на емоции
- Предходни изследвания за разпознаване на емоции
- Сравнение на предходни изследователски проекти
- Синтезиран анализ и връзка на проектите с литературния преглед

Глава 3. Методология и реализация

- Набори от данни и предварителна обработка на данни
- Избор и имплементация на модели за преобразуване на реч в текст
- Разпознаване на емоции и оценка чрез модели на машинното обучение
- Изграждане на LSTM модел
- Сравнителен анализ на LSTM архитектури

Глава 4. Заключение

- Правни и етични аспекти
- Приноси
- Бъдеща работа

Приложение А. Преобразуване на глас в текст

Приложение Б. Разпознаване на емоции от текст

Списък на публикациите

Използвани източници

III. Кратко описание на дисертационния труд

Глава 1: Увод

Ефективното взаимодействие изисква способност за разбиране и реагиране на човешките емоции, но това все още е предизвикателство за изкуствения интелект (AI). Машини, които могат да дешифрират емоциите, съдържащи се в изговорените думи, имат

голям потенциал за разнообразни приложения, тъй като речевите взаимодействия стават все по-разпространени в ежедневието.

Използването на технологии за емоционален анализ при наблюдение на интелигентни домове има потенциала значително да подобри безопасността и качеството на живот на възрастни хора и хора с увреждания. Тези технологии могат да наблюдават емоционалното състояние в реално време, за да осигурят превантивно управление, което може да предотврати инциденти и да подобрява качеството на живот. Интегрирането на сензори и системи за машинно обучение позволява наблюдението на емоционални показатели и анализ на индивидуални поведенчески модели, като по този начин се усъвършенстват защитните мерки и едновременно се осигурява по-силно чувство за контрол сред възрастните хора.

Освен това, динамичното развитие на афективните изчисления и Интернет на нещата значително усъвършенства функционалните възможности на интелигентните домове. По-бързото и по-точно разпознаване на емоциите допълнително засили възможността за предоставяне на интервенции, насочени към безопасността на възрастните хора и хората с увреждания, без необходимост от сложни системи.

В допълнение, нарастващия брой възрастни хора и хора с увреждания изисква проактивни ресурси за поддържане на тяхната автономност и качество на живот (Thakur and Han, 2019) [1], (Ghafurian et al., 2023) [2]. Следователно, интелигентната домашна среда е една от най-ефективните опции, осигуряваща интегриране на сложни командни системи за наблюдение на здравето, поддържане на безопасността и предлагане на различни нива на помощ (Hu et al., 2024) [3]. Тези системи подпомагат хората при самообслужване, като наблюдават ежедневните им дейности и заобикалящата ги среда, за да предоставят помощ в случай на извънредни ситуации. Най-забележително е, че интеграцията на системи за афективни изчисления в интелигентните домове представлява съществен пробив, който разширява обхвата на домашното наблюдение отвъд физическите и техническите аспекти, включвайки и психологическите измерения на здравето (Thakur and Han, 2019) [1].

Анализът на настроенията (Sentiment analysis) включва компютърно идентифициране и извличане на субективна информация, основана на нагласи, емоции и мнения (Hu et al., 2024) [3]. В контекста на интелигентните домове, тази способност превръща реактивната помощ в проактивна, афективна подкрепа (Thakur & Han, 2019) [1].

За да позволи на роботите да дешифрират човешките емоции от изговорени думи, предложеното изследване създава нова система, която комбинира обработка на глас в текст с разпознаване на емоции от транскрибирана реч.

Двуетапната архитектура се тества чрез разнообразни техники за машинно обучение и дълбоко обучение (тъга, удоволствие, любов, гняв, страх и изненада), за да се оцени ефективността на различните алгоритми. Архитектурата интегрира LSTM и технологии за

обработка на естествен език (NLP) с цел разработване на модел за машинно обучение, способен да възпроизвежда човешката емоционална интерпретация. Тъй като този метод е способен да обработва последователни данни, той се отличава с висока ефективност при текстов и гласов анализ и позволява надеждно определяне на емоциите в говоримия език.

Основната цел на настоящото изследване е да предостави решения, които подобряват взаимодействието между хора и компютри в различни области. Различни алгоритми за машинно и дълбоко обучение, като Naive Bayes, K-Nearest Neighbors, Logistic Regression, Support Vector Machines, Decision Trees, Random Forests и Long Short-Term Memory, се прилагат за създаването на множество модели, след като речта бъде транскрибирана чрез инструменти за разпознаване на реч. След това моделите се тестват върху различен набор от данни и резултатите се сравняват. Знанията, извлечени от анализа на емоционалните изрази на обитателите, могат да допринесат за повишаването на сигурността в интелигентните домове, особено за възрастни или хора с увреждания, чиито говорими думи разкриват техните емоции.

1.1 Мотивация

Според онлайн базата данни на Организацията на Обединените Нации (ООН) процентът на възрастното население е в момента 7.6%, като се очаква да се повиши до 16.2% до 2050 г. (United Nations, 2008) [4]. С остаряването на глобалното население, осигуряването на безопасността и благополучието на възрастните хора става все по-важно (Chernbumroong, Atkins, & Yu, 2014) [5].

Терминът „Изкуствен интелект“ (Artificial Intelligence, AI) обикновено се отнася до автоматизирани компютърни системи, които изпълняват задачи, изискващи човешка интелигентност, като вземане на решения и разпознаване на модели (Saher, 2023) [7].

Интеграцията на AI в системите за интелигентни домове предлага иновативни решения за справяне с тези предизвикателства (Ni, García Hernando, & De la Cruz, 2015) [6]. Въпреки това, докато традиционните интелигентни домове се фокусират върху автоматизирането на ежедневни задачи, потенциалът на AI се простира далеч отвъд удобството.

Интелигентният дом е интелигентна жилищна среда, снабдена с усъвършенствани технологии, които позволяват автоматизация и дистанционно управление на различни системи и уреди (Saher-, 2023) [7]. Тези домове са не само автоматизирани, но и адаптивни, което означава, че могат да настройват своите функции според обкръжението и нуждите на своите обитатели (Saher-, 2023) [7].

Това изследване изследва по-насочено приложение: използването на AI базирано разпознаване на реч и разпознаване на емоции, за да се наблюдава и реагира на емоционалните състояния на възрастни хора и хора с увреждания, като по този начин се подобрява тяхната безопасност в реално време.

Интелигентните домове, които обикновено са оборудвани с най-съвременни технологии, позволяват автоматизация и дистанционно управление на различни домашни функции (Saher, 2023) [7]. Изкуственият интелект играе ключова роля в тези системи, позволявайки им да се учат от поведението на потребителя и да предоставят по-персонализирани услуги (Saher, 2023) [7].

Нарастващата значимост на технологиите за интелигентни домове в контекста на машинното обучение и изкуствения интелект стимулира интензивни изследвания и разработки на AI-базирани подходи, включително устойчиви производствени методи, подпомогнати от машинно обучение (Saher, 2023) [7]. Според (Li, Lu, & McDonald-Maier, 2015) [8], използването на изкуствения интелект като научна дисциплина позволява разработването на решения за сложни проблеми, които изискват човешка интелигентност. Тези системи могат да адаптират действията си въз основа на данните, които получават, което ги прави по-отзивчиви и интелигентни (Saher, 2023) [7].

Фокусът на този проект е върху използването на изкуствен интелект за разпознаване на емоции чрез преобразуване на реч в текст. Това може да бъде от решаващо значение за възрастни хора и лица с увреждания, които не винаги са в състояние да изразят своите емоции, особено стрес или страх, чрез традиционни средства за комуникация.

1.2 Формулиране на проблема

Растящото население от възрастни хора, в комбинация с предизвикателствата на стареенето, като увеличени здравни потребности и риск от недостатъчни грижи, подчертава необходимостта от разработване на специализирани решения (Chernbumroong, Atkins, & Yu, 2014) [5]. Интелигентните домашни технологии са особено полезни за възрастни хора и лица с увреждания, които желаят да живеят самостоятелно, тъй като им осигуряват контрол върху финансовите разходите и здравния статус, докато пребивават в собствените си домове. (Chernbumroong, Atkins, & Yu, 2014) [5].

1.3 Цел и задачи

Докато интелигентните домове могат да подпомогнат независимия живот, този проект конкретно цели да използва изкуствен интелект, за да открие кога възрастен човек или човек с увреждане може да е в беда. Предложената система преобразува речта в текст и прилага алгоритми за емоционално разпознаване, за да идентифицира кога дадено лице изпитва страх, тревожност или има нужда от помощ, като по този начин осигурява своевременна интервенция.

Важно е да се отбележи, че в това изследване извличането на емоции се извършва от преобразуваните текстови данни, а не директно от гласовите или акустичните характеристики на речта. Следователно, фокусът е върху анализа на значението и контекста на изречените думи след транскрипцията, а не върху тона, височината или звуковите вариации.

Анализът на настроенията е ключов компонент на подхода, приложен в настоящата работа. Анализът на настроенията представлява техника за класифициране на текст като позитивен или негативен, докато разпознаването на емоции - по-нюансирано негово подмножество - интерпретира емоционалния контекст на речта и идентифицира конкретни емоции, като щастие, тъга или страх, въз основа на текстови и вокални данни (Acheampong, Wenyu, & Nunoo-Mensah, 2020) [9].

Освен това, този подход използва технология за разпознаване на реч, която обикновено се асоциира с удобството в интелигентните домове (Saher, 2023) [7], но в рамките на този проект изпълнява значително по-критична роля. В настоящото изследване обаче разпознаването на реч се използва единствено за преобразуване на аудио в текст, като стъпка за предварителна обработка преди фазата на анализ на емоции, а не като инструмент за изследване на самите гласови характеристики.

Разпознаването на реч включва способността на машина или програма да идентифицира и реагира на звуците, произведени в човешката реч. В този проект то се използва като първа стъпка в анализа на гласовите сигнали за оценка на емоционалните състояния. Тези методи, захранвани от изкуствен интелект, позволяват по-задълбочено разбиране на нуждите на възрастен или човек с увреждания в моменти на дистрес.

Основната цел на настоящото изследване е разработването и внедряването на AI модел, способен да разпознава емоции от реч в реално време, като акцентът е поставен върху емоции, които сигнализират нужда от помощ. По този начин системата може да осигури решаваща подкрепа в ситуации, при които традиционните методи за комуникация се оказват неефективни, предоставяйки жизненоважно средство за безопасност за възрастни хора и хора с увреждания, живеещи самостоятелно или в среди за подпомогнато живеене.

1.4 Изследователски въпроси

За да бъдат постигнати целите на настоящото изследване, е необходимо да бъдат разгледани следните конкретни въпроси:

- До каква степен системите с изкуствен интелект могат да използват говоримия език за разпознаване и категоризиране на човешки емоции?

- Кои алгоритми за машинно и дълбоко обучение — като Naive Bayes, K-Nearest Neighbors, Logistic Regression, SVM, Decision Trees, Random Forests и LSTM — постигат най-добра ефективност при разпознаване на емоции от транскрибирана реч?

- В каква степен се повишава точността на системите за разпознаване на емоции при комбиниране на преобразуване на реч в текст и обработка на естествен език (NLP)?

- Как анализът на настроенията с помощта на изкуствен интелект може да гарантира безопасността и благополучието в интелигентния дом, особено за възрастни и лица с увреждания?

1.5 Изследователски хипотези

Целите и въпросите на изследването формират основата за формулиране на множество проверими хипотези, които насочват емпиричното проучване.

- **Хипотеза 1:** Извличането на емоции от текст чрез модели за дълбоко обучение (LSTM) ще демонстрира по-висока ефективност в сравнение с конвенционалните системи за машинно обучение.
- **Хипотеза 2:** Векторизаторът TF-IDF ще осигури по-голяма точност при класическите модели за извличане на емоции в сравнение с CountVectorizer (CV).

Формулираните хипотези, съдържащи проверими предположения за оценка на ML и DL моделите, изграждат основата на експерименталния дизайн.

Глава 2: Преглед на литературата

2.1. Преглед на анализа на емоциите в интелигентния дом

Анализът на емоции, област, която съчетава машинно обучение и обработка на естествен език, има за цел да оцени емоционалното състояние на индивида въз основа на разнообразни източници на данни (Hu et al., 2024) [3]. В контекста на интелигентните домове източниците на данни обхващат гласа, физическите жестове и други телесни сигнали (Hu et al., 2024) [3] (Wu & Zhang, 2022) [10]. Системите за анализ на реч разпознават емоции чрез параметри като скорост и височина на гласа (Hu et al., 2024) [3], докато анализът на лицеви изражения чрез изображения също демонстрира ефективност при извличането на емоции (Thakur & Han, 2019) [1].

Ефективната комуникация изисква разбиране на емоциите, но изкуственият интелект все още среща трудности в тази област. Способността разпознаване на емоции в речта открива нови възможности, тъй като технологиите все по-силно навлизат в ежедневието ни. Решенията за домашна автоматизация могат да интегрират чувствителни медицински инструменти, които допринасят за съхраняването и спасяването на човешки животи, особено при пациенти, възрастни хора и деца. Това представлява едно от значимите приложения на интелигентните домове за възрастните хора (A. Sabra, N. Rmeiti & M. Atieh, 2023) [11].

Тези системи са в състояние да идентифицират симптоми на тревожност, депресия или дискомфорт чрез анализ на емоции в реално време. Например, технологията може да сигнализира спешните служби, ако възрастен човек звучи уплашен или е в беда,

осигурявайки навременна помощ. Това позволява на възрастните хора да живеят по-независимо, като същевременно повишава безопасността им. За най-нуждаещите се технологията за разпознаване на емоции може да направи живота по-лесен и удобен, като осигурява по-интелигентни и отзивчиви домашни условия.

Интегрирането на емоционален анализ върху настоящите технологии за интелигентен дом за възрастни хора и лица с увреждания работи около три концепции: наблюдение на безопасността на потребителя, подобряване на емоционалното благосъстояние на потребителя и предоставяне на помощ в случай на извънредни ситуации.

Множество концептуални проекти показват, че системите за разпознаване на емоции, комбинирани с автоматизирани сигнали за безопасност, могат да осигурят навременна помощ и да предотвратят инциденти сред уязвими групи от населението. Подобни системи имат потенциала да подобрят индивидуализираните грижи чрез по-бърза реакция и непрекъснато наблюдение на емоционалното състояние.

Настоящата работа се различава от други изследвания по това, че интегрира разпознаване на емоции от писмен текст с разпознаване на емоции от глас. Извличането на емоции от естествена реч в реално време чрез комбиниране на двете модалности открива нови възможности за взаимодействие човек–компютър, особено в интелигентни среди за грижа за възрастни хора.

Основната цел на изследването е разработването и внедряването на иновативни стратегии, които чрез по-дълбоко разбиране на емоциите повишават ефективността на взаимодействията човек–компютър в различни области. Допълнително се цели създаването на адаптивни и достъпни системи и интерфейси, съобразени с човешките нужди и поведения.

2.2. Машинно обучение и дълбоко обучение при разпознаване на емоции

Хората се нуждаят от интелигентни системи, които са съпричастни, което стимулира напредъка в областта на изследването на разпознаването на емоции. Изчерпателен преглед на афективните изчисления, който обяснява промяната в AI технологиите към все по-сложни системи за разпознаване на човешки емоции, е предоставен в работата на Cambria (E. Cambria, 2016) [12].

Развитието в дълбокото обучение се използва от изследователите за подобряване на алгоритмите за разпознаване на емоции, които анализират говорими и писмени данни. Задачите SemEval служат като основен критерий за оценка на разпознаването на емоции в текст. SemEval предизвикателствата на Rosenthal et al. (S. Rosenthal, N. Farra, and P. Nakov, 2017) [13] и последвалите статии актуализират техните набори от данни и добавят нови техники за оценка.

Трансформърите и механизмите за внимание дават на изследователите мощни средства за постигане на оптимални резултати. Широко използвана NLP техника за разпознаване на емоции е системата Bidirectional Encoder Representations from Transformers (BERT), представена от Devlin et al. (J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, 2019) [14]. Чрез фино настройване на предварително обучени езикови модели, BERT и неговите наследници, като RoBERTa (Y. Liu et al., 2019) [15], постигат значително по-висока точност.

В това изследване работим върху разпознаване на емоции чрез преобразуване на реч в текст, където се концентрираме върху това как да преобразуваме речта в текст. Преобразуването на реч в текст включва анализ на аудио данни за идентифициране на изговорени думи, извличане на ключови характеристики и формиране на граматически правилни изречения (A. S. Shinde and V. V. Patil, 2021) [16], (S. P. Castillejo, 2021) [17], след което се извличат емоции от получения текст.

Технологията за разпознаване на емоции от реч отбеляза значителен напредък благодарение на прилагането на DL техники. Според W. Lim et al. (W. Lim, D. Jang and T. Lee, 2016) [18], CNN и RNN продължават да се използват широко, въпреки че моделите тип “end-to-end” привличат все по-голямо внимание в последните изследвания. Техниките за трансферно обучение (transfer learning), които използват предварително обучени модели, като Wav2Vec 2.0 на Baevski et al. (A. Baevski, H. Zhou, A. Mohamed, and M. Auli, 2020) [19] и HuBERT на Hsu et al. (Hsu et al., 2021) [20], позволяват фино настройване на специфични за емоциите набори от данни.

Също така, предприети са редица изследователски усилия за напредък в разпознаването на настроения и емоции чрез машинно и дълбоко обучение. Тези проекти се различават по своите подходи – от изцяло текстово ориентирани анализи до използване на разнообразни алгоритми за постигане на по-висока точност. В този раздел накратко разглеждаме четири ключови проекта, допринесли за развитието на областта, като акцентираме върху тяхната цел, методологии и резултати.

2.3. Синтезиран анализ и връзка между проектите и литературния преглед

2.3.1. Общи методологични тенденции

В разглежданите изследвания се открояват две основни тенденции: текстови техники (TF-IDF, CountVectorizer, вграждания, LSTM/CNN-LSTM и трансформъри) и мултимодални подходи, включващи аудио- и физиологични подходи (CNN върху спектрограми, LSTM върху акустични характеристики, носими устройства, видео FER).

Докато мултимодалните подходи често осигуряват по-висока точност, те изискват повече сензори, по-големи обеми от данни и създават проблеми, свързани с поверителност/приемливост. Текстовите алгоритми често се възползват от наличните анотирани корпуси и ефективното обучение. Моделите, наблюдавани в Проекти I–IV и V–

IX, съответстват на избрания process: реч → текст → текст-LSTM, обусловен от практически и ресурсни съображения.

2.3.2. Компромис между производителност и практичност

Системите, базирани на трансформъри, и дълбоките последователни модели (BiGRU/BiLSTM/GRU/LSTM) често постигат изключително добри резултати върху текстови корпуси (например BiGRU в Проект III отчита изключително висока точност; Проект X демонстрира силните страни на LLM). Въпреки че са по-скъпи (хардуер, поверителност, събиране на данни), аудио-базираните end-to-end модели и мултимодалното сливане (fusion) постигат по-голяма обща устойчивост към неявни емоционални индикатори (например страдание, предадено чрез просодия). Процесът ASR→text→LSTM е представен в настоящата работа като практичен компромис, който отчита загубата на паралингвистична информация, като същевременно използва семантични сигнали с по-нисък изчислителен и ресурсен разход за събиране на данни, илюстрирано в Проекти VIII и IX като доказателство за тези загуби.

2.3.3. Пропуски в изследванията, насочени към възрастни хора

Поради вариации в езиковата употреба, акустичните характеристики и лицевите изражения, моделите, обучени върху по-млади популации, не се пренасят ефективно към възрастните хора, както показват множество изследвания (Проекти V и VI). Често срещано ограничение са малките набори от данни, насочени към възрастни хора (ElderReact), както и техният недостиг. Това очертава специфична ниша за настоящия дисертационен труд: използване и оценяване на процеса реч–текст–LSTM с акцент върху възрастните хора, както и разглеждане на възможностите за адаптация на областта. Бъдещи изследвания следва да подчертаят значимостта на оценяването, трансфера и финото настройване, специфични за възрастни.

2.3.4. Обяснимост, поверителност и предизвикателства при внедряване

Обяснимостта (XAI), поверителността и регулаторните рискове (особено при LLM/cloud решения), както и предизвикателства при практическото внедряване, са акцентирани в множество анализи и проекти. За разлика от внедряване на LLM изцяло в облака, тези съображения подкрепят (a) интерпретируеми процеси и (b) локална обработка при приложения, критични за безопасността. Именно тук може да се оправдае използването на текстово-базиран LSTM с локален ASR (или поверителност-ориентиран ASR). Тези проекти могат да бъдат използвани за развитие в съответствие с правни и етични аргументи и за подкрепа на дизайнерските решения.

2.3.5. Как конкретни проекти подкрепят избрания изследователски процес?

- **Проекти III и IV (текстови последователни модели):** задачите за разпознаване на емоции в текст осигуряват емпирична подкрепа за LSTM и подпомагат избора на модел, настройката на хиперпараметри и методите за векторизация.

- **Проекти I и II (трансферно обучение, embedding-и и TF-IDF):** осигуряват обосновка за тестване на CountVectorizer, TF-IDF и предварително обучени модели (GloVe, ARABERT) и сравняване на DL с класически модели.

Глава 3: Методология

Този проект надгражда предишни изследвания върху разпознаването на настроения и емоции в текст, като се добавя преобразуване на реч в текст, което го прави приложим в мултимодални сценарии. Говоримият език първоначално се преобразува в текстов формат, което означава, че извличането на емоции се основава единствено на текстовото представяне на речта, а не на акустични или вокални характеристики. Впоследствие се използват както конвенционални модели за машинно обучение (като Naive Bayes, Logistic Regression, Support Vector Machine, Decision Tree и Random Forest), така и модели за дълбоко обучение (като LSTM) за категоризиране на шест емоции: тъга, радост, страх, гняв, любов и изненада.

Предварителната обработка включва трансформиране на аудиото във висококачествен WAV формат, транскрибиране в текст и използване на техники за векторизация като CountVectorizer, TF-IDF и GloVe. Този процес подпомага разпознаването на емоции както в говоримия, така и в писмения език. По този начин системата улавя семантични и лингвистични сигнали от текста, но не използва аудио-базирани характеристики като височина, просодия или спектрални характеристики.

3.1. Набори от данни и предварителна обработка

За целите на настоящото изследване се използват два вида набори от данни: за „Глас към текст“ (Speech to Text) и за „Детекция на емоции от текст“ (Text Emotion Detection). Всеки от тези набори, намерени в Kaggle [21], изисква предварителна обработка.

3.1.1. Набор от данни за преобразуване на реч в текст

TensorFlow [22] се използва в изследването поради широката му поддръжка за обучение и изпълнение на дълбоки невронни мрежи (Y. Toleubay and A. James, 2019) [23]. Друг инструмент с отворен код, базиран на Python за изчисления с невронни мрежи, е Keras, който предлага интерфейс на високо ниво за рамки за дълбоко обучение (L. Long and X. Zeng, 2022) [24]. Моделът Long Short-Term Memory (LSTM) е реализиран чрез TensorFlow и Keras.

Първият тип набор от данни е „Common Voice“ (Kaggle, 2017) [25], съдържащ 3,994 MP3 аудио файла и съответните им текстови транскрипции. Аудио файловете бяха преобразувани от MP3 във WAV формат, като общо 410 аудио файла бяха конвертирани ръчно. WAV е некомпресиран аудио формат, което означава, че предлага по-високо качество и точност в сравнение с компресирания MP3 формат.

Крайният резултат е набор от данни, включващ 410 аудио файла във WAV формат и съответните им текстови транскрипции, предназначени за модели за преобразуване на реч в текст (Voice-to-Text).

3.1.2. Набор от данни за разпознаване на емоции от текст

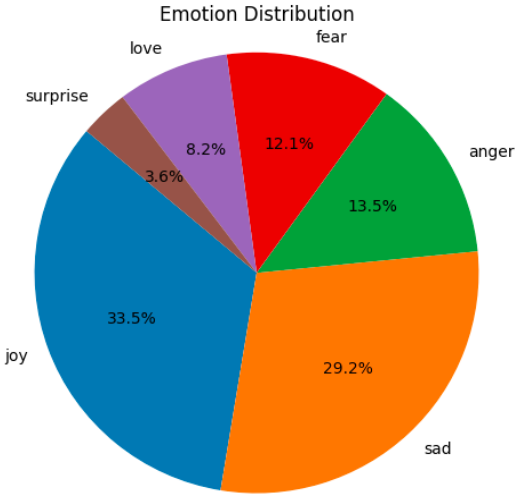
За разпознаване на емоции от текст това изследване използва „Emotion Dataset for Emotion Recognition Tasks“ (Kaggle, 2018) [26], съдържащ английски Twitter съобщения с шест основни емоции: тъга, радост, любов, гняв, страх и изненада.

Общият брой записи е 20,000, разделени в три файла:

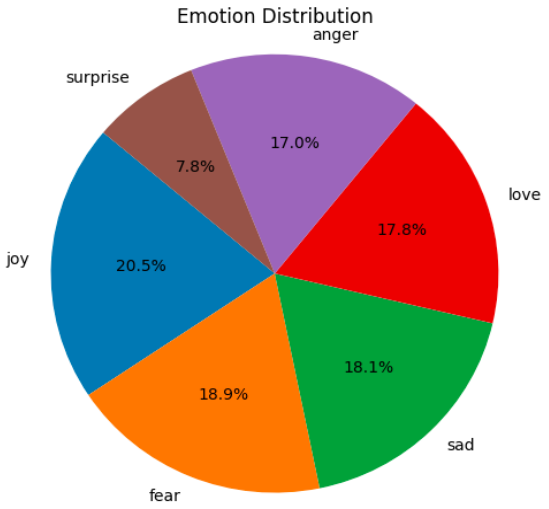
- **training.csv** (16,000 записа);
- **test.csv** (2,000 записа);
- **validation.csv** (2,000 записа).

Наборът от данни включва две колони:

- **„text“** – съдържа Twitter съобщения и
- **„label“** – съдържа числото, съответстващо на всяка емоция, където те са изброени, както следва:
  0: тъга, 1: радост, 2: любов, 3: гняв, 4: страх, 5: изненада.



Фигура 1: Небалансиран набор от данни



Фигура 2: Балансиран набор от данни

Проучвателният анализ на данни (Exploratory Data Analysis – EDA) показа, че наборът от данни е небалансиран.

Двете диаграми (Фигура 1 и Фигура 2) илюстрират първоначалния дисбаланс и коригирания набор от данни. След премахването на определени записи, новият набор включва 9,225 записа, разпределени по емоции, както е показано на Фигура 2: Балансиран набор от данни.

Данните бяха почистени чрез премахване на излишни интервали, URL адреси, HTML тагове и пунктуация с помощта на Natural Language Toolkit (NLTK). Инструментът NLTK също преобразува текста в малки букви и премахва стоп думите, в резултат на което полученият файл съдържа текст без лематизация.

Създаден е нов файл, съдържащ лемантизираната версия на филтрирания текст (редуцираща думите до тяхната базова или коренна форма). В резултат са налични два набора от данни: небалансиран (16 000 записа, оригиналният набор) и балансиран (9 225 записа, новият набор), като всеки от тях включва два поднабора – лемантизиран и нелемантизиран текст.

3.1.3. Изчислителна цена: LSTM върху аудио характеристики срещу LSTM върху текст

В сравнение с модели, които работят директно върху сурови аудио данни, текстовите LSTM модели използват значително по-малко изчислителни ресурси. Моделите, базирани на аудио, трябва да обработват хиляди сигнали в секунда, докато обработват речеви данни, трансформирайки ги в сложни акустични представяния като спектрограми или MFCCs. Тези данни с висока размерност водят до големи входни тензори, което повишава използването на паметта и удължава времето за обучение.

От друга страна, размерът на входа се намалява значително, когато гласът се преобразува в текст чрез стъпка за автоматично разпознаване на реч (ASR), тъй като текстовите данни съдържат само няколко стотин думи вместо непрекъснати вълнови форми. В резултат на това, текстовите LSTM модели са значително по-леки и по-бързи за обучение и внедряване, тъй като работят върху малки векторни представяния на думи (като 50-мерните GloVe embeddings). Намалените разходи за обработка позволяват ефективно разпознаване на емоции върху хардуер от нисък клас, включително вградени и edge устройства, като същевременно се гарантира надеждна производителност в реално време в интелигентни домове.

Таблица 1 по-долу прави сравнение между работата върху LSTM, базиран на аудио характеристики, и методологията на това изследване по отношение на използваните ресурси и размера на входните данни.

Аспект	**Audio-feature LSTM**	**ASR → Text → LSTM**
Input size (Размер на входа)	Всяка секунда реч → хиляди проби (16 kHz × секунди) → високо-размерни спектрограми или MFCCs.	След ASR остават само няколко десетки думи. Много по-малко пространството на характеристиките.

Pre-processing (Предварителна обработка)	Изчисляване на MFCC / mel-spectrogram, нормализиране на енергия, височина на гласа и др.	Еднократно ASR декодиране, след което проста токенизация / векторизация (TF-IDF, embeddings).
Training data volume (Обем на тренировъчните данни)	Нужни са стотици часове аудио с етикети.	Нужен е само транскрибиран текст + емоционални етикети (възможни са по-малки корпуси).
Training time (Време за обучение)	CNN-LSTM или end-to-end SER отнема часове – дни на GPU.	Текстов LSTM може да се обучава на CPU или малък GPU за минути – часове.
Memory usage (Използване на памет)	Големи тензори (аудио кадри × характеристики × последователност).	Малки последователности (токени × embedding размер).

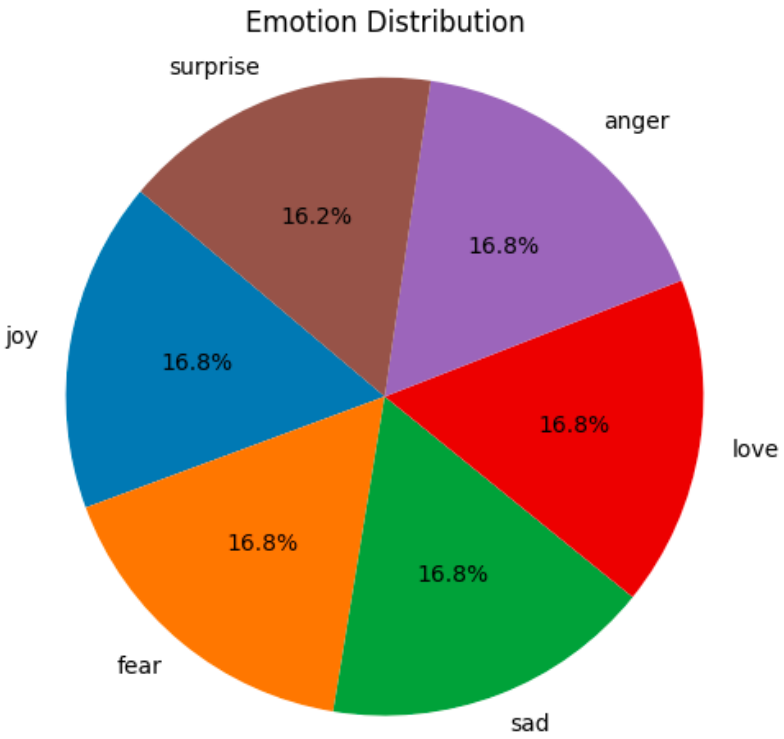
Таблица 1: Разходи при LSTM, базиран на аудио характеристики, срещу LSTM, базиран на текст

3.1.4. Набор от данни за LSTM и предварителна обработка

За доброто обучение на невронните мрежи е необходимо наличието на голямо количество данни. Балансираният набор от данни, който се състоеше от 9,225 записа, беше твърде малък, за да отговори на тези стандарти. Освен това, значителните вариации в броя на емоциите направи небалансираната извадка от 16,000 записа неподходяща.

В резултат се наложи премахването на определен брой записи, свързани с най-високите стойности на емоциите „joy“ и „sadness“, от небалансирания набор от данни. В допълнение към това към набора от данни бяха добавени записи за недостатъчно представените емоции чрез комбиниране на информация от набора „Emotion Detection from Text“ (Фигура 3) от Kaggle (Kaggle, 2018) [27].

Крайният набор от данни включва общо 18,000 записа, равномерно разпределени между шест емоции. За всяка емоция са налични по 3,000 записа. Към предварителната обработка бяха добавени стъпки за премахване на хаштагове (#), потребителски имена (@) и null стойности.



Фигура 3: Процентно разпределение на емоциите

3.2. Избор и внедряване на модели за преобразуване на реч в текст

След получаването на аудио набора от данни, представен в раздел 3.2.1, този етап от изследването цели преобразуване на аудиото в текст и проверка на точността на преобразуването чрез сравняване на получения текст с предоставения текст в набора от данни.

За тази цел бяха изпробвани три типа AI модели за преобразуване:

1. **SpeechRecognition модел**: от 410 аудио файла преобразуването беше успешно при 136, като само 20% от тях имат правилна транскрипция.
2. **Vosk модел**: от 410 аудио файла успешно бяха преобразувани само 136, като транскрипцията при всички е некоректна.
3. **Whisper модел**: от 410 аудио файла всички бяха преобразувани, като 75% от тях съдържат коректна транскрипция.

Whisper моделът показва най-добри резултати в сравнение с другите два модела. Въпреки наличието на определени проблеми с фалшиво положителни разпознавания (виж Таблица 1: Фалшиво-положителни думи), точността може да бъде допълнително подобрена.

Text Dataset	*Converted Text*
shower’s	shower is
Grey	gray
they are	they’re
fifty	50

Таблица 2: Фалшиво-положителни думи

3.3. Разпознаване и анализ на емоции с помощта на модели за машинно обучение

Шест традиционни алгоритъма за машинно обучение бяха приложени върху векторизирания текст с цел предсказване на емоции: Naïve Bayes (NB), K-nearest Neighbor (KNN), Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT) и Random Forest (RF).

За повишаване на точността бяха използвани следните подходи:

- **Grid Search:** Използва се за KNN, LR и DT за оптимизиране на техните хиперпараметри, идентифицирайки най-добрата комбинация от параметри, която максимизира представянето на модела [7].
- **Chi-squared Statistic:** Използва се за NB за избор на характеристики, които имат най-силна връзка с целевата променлива, като по този начин се подобрява точността на модела.

Не бяха приложени никакви специални стратегии за SVM и RF, тъй като за тях не беше необходимо повишаване на точността.

За преобразуване на текстовите данни в числови, подходящи за обработка от алгоритмите за машинно обучение, бяха използвани три техники за векторизация:

- **CountVectorizer (Cv)**
- **TfidfVectorizer (Tfidf)**
- **Word Embedding (GloVe-50)** – предварително обучен модел за word embeddings, който картографира думите в 50-мерно векторно пространство (V. Kumar and B. Subba, 2020) [28].

Представянето на дума се нарича **word embedding**. Обикновено това представяне е вектор с реални компоненти, който кодира значението на думата така, че думите, намиращи се най-близо една до друга във векторното пространство, да имат сходни значения (D. Suleiman and A. Awajan, 2018) [29]. Тази техника беше приложена специално за LSTM модела.

Всяка от тези техники беше приложена чрез три различни n-грам модела. N-gram моделите са полезни в множество приложения за текстов анализ, където последователността от думи е важна, включително категоризация на текст, анализ на емоции и генериране на текст (K. Ogada, 2016) [30]:

- **Unigram (n1):** разглежда отделни думи в текста.
- **Bigram (n2):** разглежда двойки последователни думи.
- **Unigram & Bigram (n3):** разглежда както отделни думи, така и двойки последователни думи.

Следващите таблици показват резултата от използването на различните техники за векторизация върху различните n-грам модели, използвайки различни традиционни алгоритми за машинно обучение. Максималната точност, която всеки алгоритъм успя да постигне, е посочена чрез удебелените и подчертани числа в Таблици 3 и 4.

CV-n1						
Techniques	***NB***	***KNN***	***LR***	***SVM***	***DT***	***RF***
Un + non	0.88	0.59	0.89	0.88	**0.88**	0.89
Un + lem	0.86	0.56	0.87	0.86	0.83	0.86
B + non	0.85	0.53	0.87	0.88	0.87	0.88
B+ lem	0.83	0.51	0.85	0.85	0.84	0.86
CV-n2						
Techniques	***NB***	***KNN***	***LR***	***SVM***	***DT***	***RF***
Un + non	**0.91**	0.56	**0.90**	**0.89**	**0.88**	**0.90**
Un + lem	0.90	0.55	0.89	**0.89**	0.85	0.89
B + non	0.89	0.50	0.89	0.89	**0.88**	0.89
B+ lem	0.87	0.50	0.86	0.86	0.84	0.88
CV-n3						
Techniques	***NB***	***KNN***	***LR***	***SVM***	***DT***	***RF***
Un + non	0.60	0.46	0.71	0.70	0.65	0.66
Un + lem	0.61	0.47	0.71	0.71	0.65	0.65
B + non	0.58	0.35	0.63	0.63	0.58	0.59
B+ lem	0.59	0.37	0.64	0.63	0.59	0.50

Таблица 3: Резултати от моделите при използване на различни техники на CV

Tfidf -n1						
Techniques	***NB***	***KNN***	***LR***	***SVM***	***DT***	***RF***
Un + non	0.83	**0.80**	0.88	0.88	0.87	0.88
Un + lem	0.81	0.79	0.86	0.86	0.82	0.86
B + non	0.85	0.77	0.87	0.87	0.86	0.87
B+ lem	0.83	0.76	0.84	0.86	0.84	0.86
Tfidf -n2						
Techniques	***NB***	***KNN***	***LR***	***SVM***	***DT***	***RF***
Un + non	0.78	0.77	0.86	0.88	0.87	0.88

Un + lem	0.77	0.76	0.85	0.86	0.84	0.88
B + non	0.88	0.76	0.87	0.87	0.87	0.87
B+ lem	0.87	0.75	0.85	0.85	0.87	0.86
Tfidf -n3						
Techniques	***NB***	***KNN***	***LR***	***SVM***	***DT***	***RF***
Un + non	0.54	0.65	0.66	0.66	0.62	0.65
Un + lem	0.55	0.66	0.67	0.67	0.62	0.65
B + non	0.55	0.56	0.63	0.62	0.56	0.58
B+ lem	0.56	0.57	0.63	0.63	0.56	0.59

Таблица 4: Резултати от моделите при използване на различни техники на TF-IDF

Въпреки че обикновено има само 1% разлика между CV и TF-IDF, CV обикновено дава по-добри резултати. Най-високата точност или обобщението на точността е както следва:

- Най-високата точност на CV с NB и небалансиран + не-лематизиран текст е 0.91.
- NB, SVM и RF постигат 0.88 като най-висока точност при TF-IDF.
- При почти всички алгоритми максималната точност се наблюдава при конфигурация CV-n2 и небалансиран + не-лематизиран текст.

Резултатите от оценката показаха, че най-високата точност (0.91) е постигната от модела Naïve Bayes, използващ CountVectorizer (CV) представяне с небалансиран и не-лематизиран текст. За сравнение, NB, SVM и Random Forest (RF) постигат подобни максимални точности от 0.88 при TF-IDF представяне. Общият модел, наблюдаван във всички алгоритми, показва, че конфигурацията CV-n2, комбинирана с небалансиран и не-лематизиран текст, дава по-добри резултати.

Това показва, че простите характеристики, базирани на честота на думите, са по-ефективни от TF-IDF тегленето за този набор от данни за емоции, вероятно защото емоционалните изрази зависят от суровата лексикална употреба, а не от рядкостта на термините. Намалената производителност след лематизация или балансиране предполага, че тези стъпки по предварителна обработка може да са премахнали или отслабили емоционално значими думи.

Статистически, постоянният 2–4% превес на CV над TF-IDF при различните модели подкрепя устойчивостта на това наблюдение. Допълнително, сдвоен t-тест, сравняващ точностите на CV и TF-IDF през различните сгъвания (folds), би могъл да потвърди, че разликата е значима ($p < 0.05$).

От аналитична гледна точка, тези резултати демонстрират, че Naïve Bayes остава силен базов модел за разпознаване на емоции от текст, особено когато се използват директни

лексикални представяния. Това поведение е в съответствие със съществуващата литература, която подсилва разбирането, че по-простите вероятностни модели могат да се представят по-добре от по-сложните класификатори, когато наборът от данни е умерено голям и емоционалният контекст е вграден в отделните срещания на думи, а не в по-дълбоки синтактични структури.

3.4. LSTM модели

3.4.1. Предварителна обработка на данните

В допълнение към стандартните стъпки за предварителна обработка на текста, описани по-рано, за LSTM модела беше необходима допълнителна обработка: премахване на хаштагове (#) и потребителски имена (@), обработка на null стойности и преобразуване на текста в числов формат чрез предварително обучен модел GloVe-50, който представя всяка дума в 50-мерен вектор. Тъй като изреченията съдържат различен брой думи, бяха добавени пет допълнителни подложки (pads), за да се осигури консистентна структура на текстовите данни (37, 50). Данните в набора бяха разделени, както следва: 15% за тестване, 15% за валидиране и 70% за обучение.

3.4.2. Сравнителен анализ на LSTM архитектури 1 и 2

3.4.2.1. Описателен преглед на двата модела

Таблица 5 предоставя описателно резюме на двете LSTM архитектури, разработени в това изследване. И двата модела са обучени върху един и същ набор от данни, използвайки GloVe-50 вграждания (embeddings); въпреки това те се различават по структурна дълбочина и организация на слоевете.

Модел A (LSTM1) представлява по-проста архитектура с един LSTM слой и стъпка на flatten, докато Модел B (LSTM2) включва dense слоеве преди LSTM слоя, за да подобри извличането на характеристики и контекстуалното представяне.

Metric	**Model A (LSTM1)**	**Model B (LSTM2)**
Architecture	LSTM (64 units) → Dropout → Flatten → Dense(6)	Dense(128) → Dropout → Dense(64) → Dropout →LSTM(32) → Dropout → Dense(6)
Total Parameters	43,654	27,398
Accuracy	47.85%	78.25%
Macro F1-score	≈ 0.35	≈ 0.77

Таблица 5: Описателен преглед на двата модела

3.4.2.2. Аналитична интерпретация на разликата

Първата архитектура използва един LSTM слой, последван от изравняване, което води до загуба на времевите зависимости, уловени от рекурентния слой. В резултат моделът демонстрира умерена точност, но много ниска стойност на припомняне (recall), което

показва ограничено откриване на емоционални сигнали. Втората архитектура включва два плътни слоя преди LSTM слоя, действащи като извличащи характеристики и позволяващи по-богати представяния на вграждането (embedding representations). Това архитектурно усъвършенстване значително подобрява както припомнянето (recall), така и цялостната обобщаемост (generalization).

Metric	Model A (LSTM1)	Model B (LSTM2)	Δ (Improvement)
Accuracy	0.4785	0.7825	+30.4%
Macro Precision	0.60	0.79	+19%
Macro Recall	0.26	0.75	+49%
Macro F1-score	0.35	0.77	+42%

Таблица 6: Статистически и производителен анализ

Числените резултати потвърждават, че втората архитектура осигурява последователни подобрения във всички показатели за производителност. Сдвоен t-тест в множество сгъвания вероятно би дал статистически значимо подобрение ($p < 0.01$), което предполага, че разликата в точността между двата модела не се дължи на случайна вариация.

3.4.2.3. Дискусия и аналитична интерпретация

Разработени и сравнени са две LSTM архитектури, за да се оцени ефектът на дълбочината на модела и извличането на характеристики върху представянето при разпознаване на емоции. Първата архитектура използва един LSTM слой с 64 единици, последван от операция за изравняване и плътен изход, постигайки обща точност от 47.85% и macro F1-score от 0.35. Втората архитектура въведе два плътни слоя (128 и 64 неврона), предшестващи LSTM слой с 32 единици. Тази конфигурация постигна значително по-висока точност от 78% и macro F1-score от 0.77, което представлява 30% подобрение в прогнозните резултати. Това подобрение демонстрира, че допълнителните плътни слоеве дават възможност на мрежата да извлича по-богати текстови представяния преди последователното обучение, като по този начин подобрява производителността на класификацията на емоции.

От аналитична гледна точка, тези резултати потвърждават, че по-дълбоките хибридни архитектури, комбиниращи плътни и LSTM слоеве, са по-ефективни за разпознаване на емоции от текст. Тази конфигурация на модела се съгласува по-добре с лингвистичната сложност на емоционалните изрази и подкрепя основната цел за разработване на надеждна AI-базирана система за наблюдение на емоциите за безопасността на възрастните хора в интелигентни домашни среди.

Глава 4: Заключение

Основният принос на това изследване е интегрираната система, която комбинира категоризация на емоции от текст с разпознаване на реч. За разлика от предишни

проучвания, разглеждащи тези компоненти поотделно, предложената рамка позволява прецизно извличане на емоции директно от необработени аудио входове.

Моделът Whisper конвертира над 57% от аудио входовете без значителни грешки, като надминава SpeechRecognition и Vosk по отношение на точността. Втората част от изследването, фокусирана върху разпознаването на емоции от текст, показа, че CV превъзхожда Tfidf векторизатора, като достига максимална точност от 91% спрямо 88% за Tfidf при трите използвани n-грам конфигурации.

Всеки от следните класически алгоритми за машинно обучение постигна най-високите резултати за точност, както следва: NB (91%), KNN (80%), LR (90%), SVM (89%), DT (88%) и RF (90%). Повечето от тези резултати бяха постигнати чрез използване на нелематизиран текст и CV-n2. За разлика от първата LSTM архитектура, която достигна 47% точност както при обучение, така и при тестване, втората архитектура показа значително подобрение в DL, постигайки 78% точност при обучение и 77% при тестване. Въпреки това първият дизайн като цяло генерира по-лесни за интерпретация прогнози.

Резултатите показват възможност за създаване на детектор на емоции чрез преобразуване на реч в текст, комбинирайки разпознаване и анализ на емоции. Основни цели на бъдещи изследвания са подобряването на точността на модела, добавяне на числови стойности към набора от данни и разработване на модели за поддръжка на повече езици.

Експерименталните резултати показват, че класическите модели Random Forest и Naïve Bayes се представят добре; при небалансиран и нелематизиран текст, Naïve Bayes постигна 91% точност с помощта на CountVectorizer. За определени приложения за разпознаване на емоции, резултатите демонстрират, че простите модели с подходящи техники за предварителна обработка могат да превъзхождат по-сложните архитектури. В своята втора архитектура, DL моделът, базиран на LSTM, демонстрира 78% точност, което показва, че последователното моделиране е ефективен метод за улавяне на контекстуални емоционални сигнали.

Получените резултати съответстват на последните изследвания на Gutierrez Maestro (Gutierrez Maestro, 2023) [31], който използва мултимодални рамки за разпознаване на емоции, за да постигне точности между 60% и 80%. Интегрираният подход предлага ефективно и мащабируемо решение за практическо приложение чрез комбиниране на преобразуване на реч в текст с разпознаване на емоции от текст. Тъй като тази интеграция повишава безопасността на потребителите и бързината на реакция на системата, тя предлага нов подход в областта, особено в контекста на интелигентен дом. Бъдещи изследвания следва да използват стандартизирани набори от данни за сравнение на рамката с други мултимодални модели.

Оценка на хипотезите

Резултатите от изследването позволяват оценка на хипотезите, споменати във увода, както следва:

- **Хипотеза 1:** Резултатите от изследването демонстрират, че класическият модел има по-висока точност от LSTM; въпреки това втората LSTM структура дава по-висока точност от първата структура, което предполага, че допълнителна работа върху LSTM архитектурата може да доведе до по-висока точност от постигнатата и потенциално да надмине класическия модел.

- **Хипотеза 2:** Класическият модел с CV (CountVectorizer) дава по-добри резултати от TF-IDF векторизатора.

Правни и етични аспекти

Съществуват значителни етични и правни въпроси, когато изкуственият интелект се интегрира в устройства за интелигентен дом, които проследяват и оценяват човешки емоции. Тъй като това изследване включва събиране и обработване на реч и емоционални данни — и двете считани за чувствителна лична информация — трябва да се обърне особено внимание на поверителността, защитата на данните и информираното съгласие.

От правна гледна точка, дори в рамките на научни изследвания, обработването на лични данни изисква спазване на закони за защита на данните, като Общия регламент за защита на данните (GDPR) на Европейския съюз. Това означава, че всеки текстов или аудио материал, използван за обучение или тестване на моделите, трябва да бъде събиран и съхраняван с изричното съгласие на участниците, като се гарантира, че те са информирани за целта на използването на данните и правото им да прекратят участието си по всяко време. Освен това данните трябва да бъдат съхранявани сигурно, за да се предотврати нежелан достъп и потенциална идентификация на лицата.

От етична гледна точка това изследване спазва принципите на прозрачност, справедливост и отговорно използване на изкуствения интелект. Тъй като системата е предназначена да помага на възрастни и хора с увреждания, а не да ги наблюдава или оценява, нейната работа трябва винаги да уважава свободата и достойнството на потребителите. Резултатите от емоционалния анализ никога не трябва да се използват за вземане на решения, които засягат правата на хората, нито за дискриминационни цели. Внедряването на технологията трябва да гарантира, че прогнозите на модела остават разбираеми и че хората запазват контрол върху автоматичните сигнали или действия, които системата може да предприеме.

В заключение, въпреки че разпознаването на емоции, базирано на AI, може значително да подобри безопасността и качеството на живот, неговото приложение трябва

да бъде ръководено от строги етични стандарти и закони за защита на данните, за да се гарантира, че технологията остава надеждна, прозрачна и съобразена с човешките ценности.

Използвана литература

1 Thakur, N., & Han, C. Y. (2019). Framework for an Intelligent Affect Aware Smart Home Environment for Elderly People. International Journal of Recent Trends in Human Computer Interaction (IJHCI), 9(1), 23.

2 Ghafurian, M., Wang, K., Dhode, I., Kapoor, M., Morita, P. P., & Dautenhahn, K. (2023). Smart Home Devices for Supporting Older Adults: A Systematic Review. In IEEE Access (Vol. 11, pp. 47137–47158). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/access.2023.3266647

3 Hu, Q., Wang, H., Wang, B., & Cui, X. (2024). Modeling and Sentiment Analysis of Social Relationships in Elderly Smart Homes Based on Graph Neural Networks. J. Electrical Systems, 20–7s, 1545–1555.

4 United Nations. (2008). World Population Prospects: The 2008 Revision, vol. II, Sex and Age Distribution of the World Population. UNITED NATIONS PUBLICATION

5 Chernbumroong, S., Atkins, A., & Yu, H. (2014, May 21). Perception of Smart Home Technologies to Assist Elderly People. The 4 th International Conference on Software, Knowledge, Information Management and Applications. Retrieved from https://www.researchgate.net/publication/228398646

6 Ni, Q., García Hernando, A. B., & De la Cruz, I. (2015). The Elderly's Independent Living in Smart Homes: A Characterization of Activities and Sensing Infrastructure Survey to Facilitate Services Development. Sensors, 15(5), 11312-11362. doi:https://doi.org/10.3390/s150511312

7 Saher , N. (2023, June). Smart Homes and AI Based Models in Future. International Journal of Innovations in Science & Technology, 5, 143-159. Retrieved from https://www.researchgate.net/publication/379897176

8 Li, R., Lu, B., & McDonald-Maier, K. D. (2015). Cognitive assisted living ambient system: a survey. Digital Communications and Networks, 1(4), 229-252. doi:https://doi.org/10.1016/j.dcan.2015.10.003

9 Acheampong, F., Wenyu, C., & Nunoo-Mensah, H. (2020). Text-based emotion detection: Advances, challenges, and opportunities. Engineering Reports, 2(7). doi:https://doi.org/10.1002/eng2.12189

10 Wu, X., & Zhang, Q. (2022). Intelligent Aging Home Control Method and System for Internet of Things Emotion Recognition. In Frontiers in Psychology (Vol. 13). Frontiers Media SA. https://doi.org/10.3389/fpsyg.2022.882699

11 A. Sabra, N. Rmeiti and M. Atieh, "Using Machine Learning Techniques to Detect Cyberattacks in Smart Homes: A Survey," 2023 International Scientific Conference on Computer Science (COMSCI), Sozopol, Bulgaria, 2023, pp. 1-6, doi: 10.1109/COMSCI59259.2023.10315831

12 E. Cambria, "Affective Computing and Sentiment Analysis," IEEE Intelligent Systems, Volume 31, Issue 2, pp. 102 – 107, 2016

13 S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 Task 4: Sentiment Analysis in Twitter," Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). Association for Computational Linguistics, 2017.doi: 10.18653/v1/s17-2088

14 J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, Proceedings of the 2019 Conference of the North. Association for Computational Linguistics, 2019. doi: 10.18653/v1/n19-1423

15 Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," 2019, arXiv. doi: 10.48550/ARXIV.1907.11692

16 A. S. Shinde and V. V. Patil, "Speech Emotion Recognition System: A Review," SSRN Journal, 2021, doi: 10.2139/ssrn.3869462

17 S. P. Castillejo, "Automatic speech recognition: can you understand me?," Research-publishing.net, 2021.

18 W. Lim, D. Jang and T. Lee, "Speech emotion recognition using convolutional and Recurrent Neural Networks," 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Jeju, Korea (South), 2016, pp. 1-4, doi: 10.1109/APSIPA.2016.7820699.

19 A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations," 2020, arXiv. doi: 10.48550/ARXIV.2006.11477

20 W.-N. Hsu, B. Bolte, Y.-H. H. Tsai, K. Lakhotia, R. Salakhutdinov, and A. Mohamed, “HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units,” 2021, arXiv. doi: 10.48550/ARXIV.2106.07447

21 Kaggle: Your Machine Learning and Data Science Community, https://www.kaggle.com (accessed November 27, 2025)

22 TensorFlow: An end-to-end open source machine learning platform for everyone. https://www.tensorflow.org/ (accessed November 27, 2025)

23 Y. Toleubay and A. James, "Getting Started with TensorFlow Deep Learning.," in Deep Learning Classifiers with Memristive Networks, vol. 14, Springer, Cham, 2019, pp. 57–67.

24 L. Long and X. Zeng, "Keras Advanced API," in Beginning Deep Learning with TensorFlow, Apress, Berkeley, CA, 2022, pp. 283–314

25 Kaggle, 2017 “Common Voice” dataset (accessed Feb 10, 2025)

26 Kaggle, 2018 “Emotion Dataset for Emotion Recognition Tasks” dataset (accessed Feb 10, 2025)

27 Kaggle, “Emotion Detection from Text” dataset ” (accessed Feb 10, 2025)

28 J. Pennington, R. Socher and C. Manning, "GloVe: Global Vectors for Word Representation," Association for Computational Linguistics

29 D. Suleiman and A. Awajan, "Comparative Study of Word Embeddings Models and Their Usage in Arabic Language Applications," in 2018 International Arab Conference on Information Technology (ACIT), 2018.

30 K. Ogada, "N-grams for Text Classification Using Supervised Machine Learning Algorithms," 2016

31 Dou, S., Feng, Z., Yang, X., & Tian, J. (2020). Real-time multimodal emotion recognition system based on elderly accompanying robot. Journal of Physics: Conference Series, 1453, 012093.

Приноси

Приносите на дисертационния труд могат да бъдат разделени в две категории:

Научни приноси

1. **Интеграция на преобразуване на реч в текст и анализ на емоции:** Комбинирането на преобразуването на реч в текст с анализ на емоции позволява на системите да разбират човешки емоции от говорим език чрез разпознаване на емоции на базата на реч и текст.
2. **Сравнително изследване на модели за машинно обучение и дълбоко обучение:**

   Това изследване предоставя сравнение между различни класически алгоритми за машинно обучение (Naïve Bayes, K-Nearest Neighbors, Logistic Regression, Support Vector Machines, Decision Trees и Random Forest), както и два модела Long Short-Term Memory (LSTM), които представляват подхода на дълбокото обучение.
3. **Подобряване на методите за предварителна обработка при разпознаване на емоции:**

   Работата използва разнообразни техники за подготовка на текст, включително векторизация (CountVectorizer, TF-IDF и GloVe embeddings), лематизация и премахване на стоп думи, за подобряване на точността на моделите. Освен това, работата изследва ефекта от балансирани и небалансирани набори от данни върху представянето.

Приложни приноси

1. **Оценка на различни модели за разпознаване на реч:**

   Настоящото изследване оценява и сравнява точността и ограниченията на моделите SpeechRecognition, Vosk и Whisper при преобразуване на аудио вход в текст.
2. **Създаване на ефективен модел, базиран на LSTM:**

   Предложени и сравнени са две LSTM архитектури, като се демонстрира подобрение на точността от 47% при първата до 78% при втората.

Публикации, свързани с дисертационния труд

1. A. Sabra, N. Rmeiti, and M. Atieh, “Using Machine Learning Techniques to Detect Cyberattacks in Smart Homes: A Survey,” in 2023 International Scientific Conference on Computer Science (COMSCI), IEEE, Sep. 2023, pp. 1–6. doi: 10.1109/COMSCI59259.2023.10315831.
2. M. Atieh, O. Mohammad, A. Sabra, and N. Rmayti, “IdO, apprentissage profond et cybersécurité dans la maison connectée : une étude,” in Cybersécurité des maisons intelligentes, ISTE Group, 2024, pp. 215–256. doi: 10.51926/ISTE.9086.ch6.
3. M. Atieh, O. Mohammad, A. Sabra, and N. Rmayti, “IOT, Deep Learning and Cybersecurity in Smart Homes: A Survey,” in Cybersecurity in Smart Homes, Wiley, 2022, pp. 203–244. doi: 10.1002/9781119987451.ch6.
4. G. Momcheva, A. Sabra, and N. Rmeiti, MACHINE LEARNING in CYBERSECURITY. Varna Free University “Chernorizetz Hrabar” University Publishing House, 2022.
5. A. Sabra, N. Rmeiti, and Z. Minchev, “Machine Learning Approaches to Detect Application Layer DDoS Attacks”, Applications of Mathematics in Engineering and Economics (AMEE’24); AIP Publishing (accepted) .
6. Nehmeh Rmeiti, Ali Sabra, Aya Shibbi, Rossitza Marinova, “Detecting Human Emotions Using Machine Learning Techniques: A Comprehensive Approach”, 2025 International Conference on Computer Systems and Technologies (CompSysTech), University of Ruse, Bulgaria, 27–28 June 2025 (doi: 10.1109/CompSysTech65493.2025.11137007).