



**ТЕХНИЧЕСКИ УНИВЕРСИТЕТ – СОФИЯ**  
**Факултет „Компютърни система и технологии“**  
**Катедра „Информационни технологии в индустрията“**

---

Маг. инж. ХРИСТИНА ПЕТРОВА НИКОВА

**СЪВРЕМЕННИ МЕТОДИ ЗА ОБРАБОТКА И АНАЛИЗ НА  
ИНФОРМАЦИЯ ОТ ИНФОРМАЦИОННО-ИЗМЕРВАТЕЛНИ  
СИСТЕМИ**

**АВТОРЕФЕРАТ**

На дисертация за получаване на образователна и научна степен  
**„ДОКТОР“**

Област: 5. Технически науки

Професионално направление: 5.3. Комуникационна и компютърна техника

Научна специалност: Автоматизирани системи за обработка на  
информация и управление

Научни ръководители:

проф. д-р инж. Ташо Ангелов Ташев

доц. д-р инж. Радослав Цветков Делийски

София

2024

Дисертационният труд е обсъден и насочен за защита от Катедрения съвет на катедра „Информационни технологии в индустрията“ към факултет „Компютърни система и технологии“ на ТУ-София на редовно заседание, проведено на 09.04.2024 г.

Публичната защита на дисертационния труд ще се състои на 08.07.2024 г. от 13 часа в Конферентната зала на БИЦ на Технически университет – София на открито заседание на научното жури, определено със заповед ОЖ-5.3-18/18.04.2024 г. на Ректора на ТУ-София в състав:

1. проф.д-р Румен Иванов Трифонов
2. доц.д-р Радослав Цветков Делийски
3. проф.д-тн Петко Христов Петков
4. проф.д-р Атанас Велков Атанасов
5. доц.д-р Стефан Михайлов Филипов

Рецензенти:

1. проф.д-р Румен Иванов Трифонов
2. проф.д-р Атанас Велков Атанасов

Материалите по защитата са на разположение на интересувашите се в канцеларията на факултет „Компютърни система и технологии“ на ТУ-София, блок № 1, кабинет 1443А.

Дисертантът е редовен докторант към катедра „Информационни технологии в индустрията“ на факултет „Компютърни система и технологии“. Изследванията по дисертационната разработка са направени от автора.

Автор: маг. инж. Христина Петрова Никова

Заглавие: Съвременни методи за обработка и анализ на информация от информационно-измервателни системи

Тираж: 30 броя

Отпечатано в ИПК на Технически университет – София

## **I. ОБЩА ХАРАКТЕРИСТИКА НА ДИСЕРТАЦИОННИЯ ТРУД**

### **Актуалност на проблема**

Информационно-измервателните системи (ИИС) са основно средство за регистриране и обработване на информация в различни сфери и конкретно за постигане на ефективен мониторинг, прогнозиране, превенция и ранно локализиране на горските пожари. Въпреки забележимите достижения в развитието на технологиите през последните години, все още съществуват значителни затруднения по отношение на намиране на ефективно решение за прогнозирането на появата на горски пожари.

Тенденциите през последното десетилетие отчитат горски пожари с необичайно големи размери и обхват. Нанесените щети поставят рекордни нива в историята досега. Според изследвания на ООН в бъдещ план се прогнозира значително увеличение на случаите на горски пожари в световен мащаб - съответно с 14% до края на 2030 и с 30% до края на 2050.

В дисертацията като се използват съвременни методи за обработка и анализ на информация от информационно-измервателни системи, са предложени модели, обещаващи добри перспективи за повишаване степента на точност на прогнозирането на вероятността от възникване на горски пожари.

### **Цел на дисертационния труд, основни задачи и методи за изследване**

На база на направените анализи, се поставя следната цел на дисертационния труд: Разработване и изследване на модели за прогнозиране на вероятността от горски пожари базирани на информация от реална информационно-измервателна система.

За постигането на целта ще бъдат изпълнени следните конкретни задачи:

1. Разработване на експериментален статистически модел за определяне на вероятността за възникване на горски пожари и определяне влиянието на основните променливи влияещи върху вероятността;
2. Анализ и оценка на адекватността на разработения статистически модел;
3. Разработване и валидиране на експериментален интелигентен модел за прогнозиране на вероятността от възникване на горски пожар чрез използването на многослойна изкуствена невронна мрежа (ANN).
4. Извършване на сравнителен анализ между разработените модели и общоприетите индекси за пожарна опасност за оценка за адекватност им.

### **Научна новост**

Разработен е експериментален статистическият нелинеен модел на бинарната логистична регресия за определяне на вероятността от възникване на горски пожар, базиран на информация от реална информационно-измервателна система. Моделът е разработен за особеностите на района на Уитмор, Северна Калифорния, САЩ. Изследвана е взаимовръзката между 12 независими променливи, представляващи метеорологични и топографски параметри, и наличието на пожар, като статистически е оценена значимостта на всяка от независимите променливите.

Предложен е общ алгоритъм за разработване и валидиране на многослоен изкуствено-невронен предиктор за прогнозиране на горски пожари, който е разработен за района на Уитмор, но може да се приложи при анализиране на опасността от пожар за различни географски региони с различни входни данни. ANN предикторът е с висока степен на точност и осигурява бърза обработка на данните, поради използваната проста архитектура и бързо сходящият метод за обучение.

Разработката по тематиката на дисертационната работа представлява научна новост. Не е известно да има образци на разработените модели, работещи по предложения начин с използваните подходи и данни за района на Уитмор.

### **Практическа приложимост**

Получените резултати от дисертационния труд са с теоретична и практическа насоченост към проблемите относно прогнозирането на вероятността за възникване на горски пожари и са свързани с опазването на околната среда и предотвратяването на последствията от потенциална екологична криза. Чрез използването на съвременни методи за обработка и анализ на информация от реални информационно-измервателни системи, са разработени експериментални модели за прогнозиране на вероятността за възникване на горски пожари с висока степен на точност.

Създаден е общ алгоритъм, базиран на принципите на изкуствения интелект, който може да бъде приложен при анализиране на опасността от пожар за различни географски региони с различни входни данни.

С практическата реализация на дисертационните резултати се повишава ефективността на системите за мониторинг и контрол на горски пожари по света.

### **Апробация**

Резултатите от изследванията, публикувани в настоящия дисертационен труд, са получили обективна оценка на изнесени доклади и презентации при участие на международни научни конференции и публикация в индексирано научно списание.

### **Публикации**

Отпечатани са 4 публикации, от които една самостоятелна и 3 в съавторство. Сред тях две са индексирани в SCOPUS, в това число една статия в списание „WSEAS Transactions on Environment and Development” и една в сборник с доклади от „XI International Scientific Conference on Computer Science (COMSCI)” под егидата на IEEE. Две публикации са отпечатани в сборник с доклади от 19-та международна конференция „Challenges in Higher Education and research in the 21th century”.

### **Структура и обем на дисертационния труд**

Дисертационният труд е в обем от 133 страници, като включва увод, четири глави за решаване на формулираните изследователски задачи, списък на приносите, списък на научните публикации, свързани с дисертационния труд и използвана литература. Цитирани са общо 120 литературни източници. Работата включва 31 фигури и 15 таблици. Номерата на литературните източници, фигурите и таблиците в автореферата съответстват на тези в дисертационния труд.

### **БЛАГОДАРНОСТИ**

*Авторът изказва своите благодарности към проф. д-р инж. Ташо Ташев и доц. д-р инж. Радослав Делийски за научно-методичното ръководство на дисертационния труд., на проф. д-р инж. Снежана Йорданова за ценните съвети и препоръки при подготовка на дисертацията, както и на д-р Faith Ann Heinsch от Департамента по научноизследователска дейност към Министерството на земеделието и горите, Монтана, САЩ за предоставената информация и насоки за работа.*

*С огромна признателност и благодарност!*

## **II. СЪДЪРЖАНИЕ НА ДИСЕРТАЦИОННИЯ ТРУД**

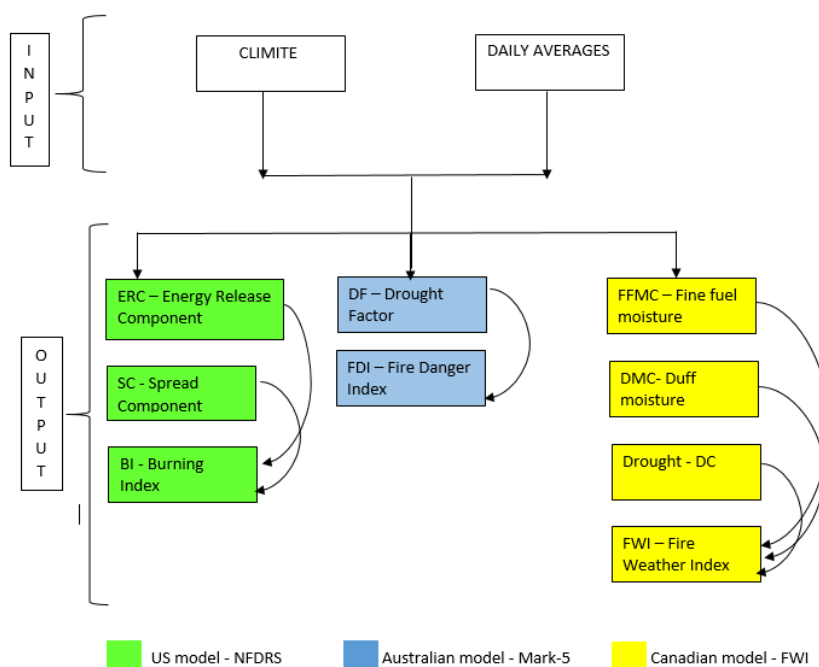
## ГЛАВА I.

### ОБЗОР НА МЕТОДИТЕ И СИСТЕМИТЕ ЗА АНАЛИЗ НА ИЗМЕРВАТЕЛНА ИНФОРМАЦИЯ ЗА ОЦЕНКА НА ОПАСНОСТТА ОТ ГОРСКИ ПОЖАРИ

#### 1.4.1. Методики за мониторинг и прогнозиране на горски пожари

В световен мащаб са разработени различни методики за ранно откриване, предизвестяване и предотвратяване на горски пожари. Три от най-разпространените методики за оценка на опасността от горски пожари, имащи голям принос в областта на изследването на възникването на пожари, са Националната методика за оценка на пожарната опасност на САЩ - U.S. National Fire Danger Rating System (NFDRS) [49][50], Канадската методика за оценка на индекса за пожароопасност на горските служби - Canadian Forest Service Fire Weather Index Rating System (FWI) [51][52] и Австралийската рейтингова методика McArthur Mark 5: Fire danger index (FDI) [53][54].

Всяка от трите методики представя, като изходен резултат от направените изследвания, съвкупност от индекси, които характеризират различни условия на възникване на горски пожар (Фиг.1.12).



Фиг. 1.12. Схема на основните компоненти на трите методики за оценка на опасността от горски пожари (NFDRS, Mark 5 и FWI).

В САЩ NFDRS разработват и използват три основни индекса - Spread Component (SC), Energy Release Component (ERC) и Burning Index (BI). В Mark 5 основните индекси, въз основа на които се прави прогнозата за възпламеняване на горски пожар, са Drought factor (DF) и Fire Danger index (FDI). В канадската FWI има три кода - Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC) – които съставят окончателния индекс за опасност от пожар - Fire Weather Index (FWI). [49][55-59]

NFDRS е една от най-широкоразпространените рейтингови противопожарни методики. NFDRS предоставя като резултат изходни индекси, изчислени на всеки кръгъл час за 24 часа в денонощието. Всеки един от тях характеризира в различен аспект възникването на пожар [49].

В следващата точка са анализирани детайлно индексите за оценка на опасността от горски пожари към американската противопожарна методика NFDRS, които са използвани за сравнителен анализ с разработените в настоящия труд модели. Сравнителният анализ е представен в Глава IV.

#### **1.4.2. Аналитични методи за прогнозиране на горски пожари.**

##### **○ Индекс на разпространение на огъня - Spread Component (SC)**

Spread Component (SC) е компонент, който определя разпространението на огъня. Индексът зависи от три фактора - финия горим материал (FM1), скорост на вятъра и наклона на терена. Тежките горими материали (FM10, FM100, FM1000) не се включват в изчислението на SC. Поради зависимостта си от финия горим материал (FM1) и скоростта на вятъра, стойностите на SC може да варират значително ежедневно. SC отговаря на въпроса „Колко бързо?“ се разпространява пожара. Той е индекс за оценка на скоростта на разпространение на огъня [49][55].

Математическата формула за изчисляване на индекса SC е следната:

$$SC = R \cdot (1 + \Phi_w + \Phi_s) \quad (1.12)$$

##### **○ Индекс на топлинна енергия - Energy Release Component (ERC)**

Индексът ERC (Energy Release Component) се отнася до наличната топлинна енергия на единица площ, разглеждана във фронталната зона на разпространение на пламъка [49][59].

Математическият израз за ERC е както следва:

$$ERC = 0,04 \cdot H_a = 0,04 \cdot I_R \cdot t_R, \quad (1.13)$$

където:

- $H_a$  – топлинна енергия на единица площ, [BTU/ft<sup>2</sup>]
- $I_R$  – реакцията на интензитета, [BTU/ FT<sup>2</sup>/min]
- $t_R$  – време, за което дадена зона е под действието на фронталния огън, [min]

ERC зависи от „тежките“ горими материали (FM100 и FM1000), което го определя като добър показател за отчитане на сезонни промени. За разлика от SC, ERC не зависи от факторите вятър ( $\Phi_w$ ) и наклон на терена ( $\Phi_s$ ), следователно ежедневната промяна в стойностите на индекса, ще бъде относително малка и ще се наблюдава сравнително рядко. ERC е индикатор за поведението на огъня като следствие от процеса на засушаване в среден до дългосрочен план [64].

#### ○ Индекс на изгаряне (BI)

Индексът на изгаряне (BI) е компонент, който оценява опасността от пожар, свързана с потенциалната дължина на пламъка в дадена областта. BI се изразява като числова стойност, равна на стойността на дължината на пламъка, изразена като мерна единица във футове, умножена по 10.

Индексът на изгаряне (BI) е произведение от компонентата на разпространение на пожара (SC) и компонентата за освобождаване на енергия (ERC). Той комбинира бързата реакция на индекса SC с по-бавния отговор на индекса ERC. Поради зависимостта на SC от скоростта на вятъра и финия горим материал (FM1), стойностите на BI варират всекидневно [49][55][59]. Индексът на изгаряне (BI) отговаря на въпроса „Колко трудно?“ би се овладял даден пожар.

Крайната формула за изчисляване на BI е изразена със следния математически израз:

$$BI = \left( \frac{ERC * SC}{0.091} \right)^{0.046} \quad (1.14)$$

Трите индекса са с отворена горна граница. Стойностите се основават, както на измерванията от метеорологични станции, така и на типа на района, за който се прави анализа. Трите индекса отразяват различни характеристики на опасността от възникване на горски пожар.

С цел оптимизация на процесите на прогнозиране, предотвратяване и противодействие на горски пожари, е разработена концепцията за комбиниране на два от индексите - ERC и BI в една метрична единица и е формулирана в нов индекс, наречен Severe Fire Danger Index (SFDI) [65-68].



### ○ Severe Fire Danger Index (SFDI)

Обединявайки тези два основни индекса, се комбинират две противоположни, но и потенциално пожароопасни условия - „сухо и спокойно“ състояние и „влажно и ветровито“ състояние. [65]. Докато първото условие, което дефинира високи стойности на ERC и ниски на BI, води до отделяне на големи количества топлина, то второто условие, което предполага ниски стойности на ERC и високи на BI, води до бързи скорости на разпространение на пожара и голяма дължина на пламъка. Следователно, съпоставянето на високи стойности на ERC с високи стойности на BI трябва да определя местата с най-голям потенциал за екстремни пожароопасни събития.

В резултат на това е представен по-обобщен параметър - индексът Severe Fire Danger Index (SFDI), който е произведение от нормализираните стойности на ERC и BI:

$$SFDI' = (ERC' * BI)'$$
 (1. 15)

### 1.5. Изводи

1. Направен е преглед на съществуващите съвременни методи за анализ на информация от информационно-измервателни системи, като е обърнато внимание на методите за предсказване на бъдещи събития базирани на реално измерени данни.

2. Информационно-измервателните системи играят съществена роля за осигуряване на информация необходима за прогнозиране и ранно известяване на горски пожари.

3. Показана е класификация и систематизация на основните методики и индекси за мониторинг, оценка и прогнозиране на степента на пожарна опасност в горските райони. Разгледани са основните параметри, оказващи влияние на възникването на горски пожари подлежащи на измерване.

4. Увеличаването на честота и интензитета на горските пожари през последните десетилетия, съпроводено от все по-неефективното им прогнозиране превръща масово използваните подходи, базирани на модела за повърхностно разпространение на пожара на Ротермел в неефективни. Поради тази причина необходимостта от изграждане на модел с висока степен на точност, се очертава като задача с приоритетно значение.

## ГЛАВА II.

### РАЗРАБОТВАНЕ НА СТАТИСТИЧЕСКИ МОДЕЛ ЗА АНАЛИЗ И ПРОЗНОЗИРАНЕ НА ВЕРОЯТНОСТТА ОТ ВЪЗНИКВАНЕ НА ГОРСКИ ПОЖАРИ

#### 2.1. Въведение

За изпълнение на задачата за разработване на статистически модел за прогнозиране на вероятността от възникване на горски пожари са събрани данни от Националният изследователския център за оценка на опасността от горски пожари в Съединените Американски щати – The National Fire Danger Rating System (NFDRS).

Необходимо е да се анализират, моделират и оценят възможните функционални зависимости между метеорологичните условия и типа на растителността за даден регион, от една страна, и възникването на горски пожари, от друга. Освен това е необходимо и да се определи типа и значимостта на тези причинно-следствени връзки, въз основа на което да се изгради модел за предсказване вероятността (P) от възникване на пожар.

С помощта на статистическата софтуерна платформа на IBM® - SPSS® Statistics 29 [70] ще бъде изграден модел, в който като входни независими променливи (X) са заложили параметрите, характеризиращи метеорологични условия. Независими променливи (X) са наречани още предиктори на зависимата променлива (Y). За изходната зависимата променлива (Y) имаме една двоична величина, заемаща само 2 състояния - „наличие на пожар“ или „липса на пожар“.

За най-подходящ статистически метод за отразяване на функционалните зависимости между атмосферните особености и възникването на пожар е избран методът на бинарната логистична регресия. Този метод съответства най-точно на заложените в изследването цели за откриване и оценка на причинно-следствена връзка между входните и изходните данни на разглеждания модел, тъй като е базиран на дихотомния тип на изходната зависимата променлива (Y).

За генерирането на модела е използвана извадката, включваща 20 от най-големите по площ горски пожари за 10-годишен период от 2007 до 2017 г. в района на Уитмор (Whitmore), Северна Калифорния, САЩ. Моделът е разработен за конкретния район на Уитмор, отчитайки географските и климатични особености региона.

## 2.4. Теоретична постановка на процеса на моделиране на бинарна (двоична) логистична регресия

Изобразена графично функцията на логистичната регресия, с величина по ординатната ос вероятността (P), има S-образна форма на разпределение, която наподобява функцията на нормално разпределение, но с ограничение на стойностите на прогнозираните вероятности (P) в диапазона между 0 и 1. Логистичното разпределение се дефинира с логистичната (сигмоидната) функция, която се описва със следния математически израз:

$$F(g(x)) = P = \frac{1}{1+e^{-y}} = \frac{e^y}{1+e^y} = \frac{e^{(\beta_0+\beta_1x)}}{1+e^{(\beta_0+\beta_1x)}} \quad (2.1)$$

За определяне стойностите на параметрите на функцията е използван методът на максималното правдоподобие (Maximum Likelihood Estimation). Тук търсените елементи са стойностите на тегловите коефициентите  $\beta_1, \beta_2 \dots \beta_k$  на съответните независими променливи  $x_1, x_2 \dots x_k$ . За категория „наличие на пожар“, означена с „1“, изчисляването на коефициентите  $\beta$  е по такъв начин, че произведението от всички вероятности "p(xi)" да клони като стойност към единица. При категория „липса на пожар“, означена с „0“, коефициентът се изчислява, така че произведението от всички вероятности "1 - p(xi)" да клони към нула. Математическият израз на гореописаното е следният:

- Продукт от всички стойности на вероятност при събитие „наличие пожар“  
 $[p(xi)] : \prod_{s \text{ in } y_i=1} p(xi) \quad (2.11)$

- Продукт от всички стойности на вероятност при събитие „липса на пожар“ [1 - p(xi)] :  $\prod_{s \text{ in } y_i=0} (1 - p(xi)) \quad (2.12)$

Така получаваме функцията на правдоподобие (likelihood function) L( $\beta$ ) като произведение от продуктите на вероятностите на двете събития, която има следния вид:

$$L(\beta) = \prod_{s \text{ in } y_i=1} p(xi) * \prod_{s \text{ in } y_i=0} (1 - p(xi)) \quad (2.13)$$

Получената функция трябва да се оптимизира за намиране на стойностите на коефициентите  $\beta$  при максимални стойности на произведението. Произведението от (2.13) представено в логаритмичен вид е както следва:

$$L(\beta) = \prod_s (p(xi)^{y_i} * (1 - p(xi))^{1-y_i}) \quad (2.14)$$

$$l(\beta) = \sum_{i=1}^n y_i \log(p(xi)) + (1 - y_i) \log(1 - p(xi)) \quad (2.15)$$

Заместваме  $p(x_i)$  с експоненциалната му форма  $p(x_i) = \frac{1}{1+e^{-\beta x_i}} = \frac{e^{\beta x_i}}{1+e^{\beta x_i}}$  и

получаваме:

$$l(\beta) = \sum_{i=1}^n y_i \log\left(\frac{1}{1+e^{-\beta x_i}}\right) + (1 - y_i) \log\left(\frac{e^{-\beta x_i}}{1+e^{-\beta x_i}}\right) \quad (2.16)$$

$$l(\beta) = \sum_{i=1}^n y_i \left[ \log\left(\frac{1}{1+e^{-\beta x_i}}\right) - \log\left(\frac{e^{-\beta x_i}}{1+e^{-\beta x_i}}\right) \right] - \log\left(\frac{e^{-\beta x_i}}{1+e^{-\beta x_i}}\right) \quad (2.17)$$

$$l(\beta) = \sum_{i=1}^n y_i [\log(e^{\beta x_i})] + \log\left(\frac{e^{-\beta x_i}}{1+e^{-\beta x_i}} * \frac{e^{\beta x_i}}{e^{\beta x_i}}\right) \quad (2.18)$$

$$l(\beta) = \sum_{i=1}^n y_i \beta x_i + \log\left(\frac{1}{1+e^{\beta x_i}}\right) \quad (2.19)$$

Следователно получаваме крайния математическият израз за определяне функцията на правдоподобие, който има следния вид:

$$l(\beta) = \sum_{i=1}^n y_i \beta x_i - \log(1 + e^{\beta x_i}) \quad (2.20)$$

#### 2.4.2. Методи за намиране на максимума на функцията на правдоподобие (Методи за максимално правдоподобие- Maximizing Log-likelihood function - MLE)

След извеждането на функцията на правдоподобие, целта е да се намерят най-правдоподобните стойности на тегловите коефициенти  $\beta$ , за които функцията на правдоподобие да приема максимална стойност.

Методите за максималното правдоподобие са техниките за намиране на максимума на функцията на правдоподобие  $l(\beta)$ . Те представляват числени методи за нелинейни изрази. Най-често използваните методи за максималното правдоподобие са: Методът на бисекцията, Методът на Мюлер, Методът на Нютон-Рафсън, сред които най-широко приложим е методът на Нютон-Рафсън.

Методът на Нютон-Рафсън е подходът, който се прилага за намиране на максимума на функцията на правдоподобие на разработвания експериментален математически модел.

### 2.3 Входни данни. Процес по събиране на необходимите данни за изграждане на модел

В използваната извадка от 20 пожара, като входни данни за всеки един от тях, са включени 10 параметъра, характеризиращи атмосферните условия и типове на горимите природни материали за дадения район. Избраните параметри, които имат най-голямо влияние върху потенциалната опасност от възникване на пожар са: атмосферна температура (Temp), относителна влажност на въздуха (RH), слънчева радиация (SolR),

количество валежи (Rain), скорост на вятъра (Wind), минимална температура за деня (MinT), максимална температура за деня (MaxT) и 5 индекса, описващи типа на горимите природни материали и водното съдържание в тях (FM1, FM10, F100, FM1000, KBDI).

Входните данни, представляващи независимите променливи, се регистрират на всеки кръгал час 24 часа в денонощието. Докато за изходни данни имаме една зависима променлива от биномен тип, описваща две състояния - на наличие на пожар (Fire) или липса на пожар (No fire).

В табл. 2.2. са посочени избраните 20 пожара, съответните дати и часове на възникване, класовете, размерите и географското местоположението на всеки пожар.

Табл. 2.2. Характеристика на пожарите, използвани за разработване на модела

<b>FIRE</b>	<b>FIRE_SIZE</b> <b>[ha]</b>	<b>FIRE_SIZE</b> <b>CLASS</b>	<b>LATITUDE</b>	<b>LONGITUDE</b>	<b>DATE</b>	<b>Time</b>
1	9356	G	40,683	-121,467	1 Aug 2009	12:00
2	2054	F	40,623	-121,955	21 Jun 2008	13:00
3	20	C	40,621	-122,012	20 Aug 2011	17:00
4	557	E	40,724	-122,023	22 Jun 2008	13:00
5	47958	G	40,719	-121,552	31 Jun 2014	16:00
6	95	C	40,631	-122,031	25 Sep 2016	7:00
7	996	E	40,976	-121,825	12 Sep 2017	14:00
8	78	C	40,677	-121,995	2 Aug 2015	3:00
9	27676	G	40,383	-121,835	18 Aug 2012	10:00
10	28336	G	40,717	-122,434	23 Jun 2008	15:00
11	8056	G	40,492	-122,499	9 Sep 2013	13:00
12	710	E	40,768	-122,601	24 Aug 2016	14:00
13	343	E	40,846	-122,365	07 Sep 2008	13:00
14	28079	G	40,495	-121,454	26 Jun 2011	17:00
15	988	E	40,834	-122,319	01 Aug 2012	14:00
16	241	D	40,814	-122,321	21 Sep 2012	16:00
17	28079	G	40,495	-121,454	23 Jul 2012	17:00
18	215	D	40,377	-122,728	09 Jul 2013	12:00
19	1375	F	40,730	-122,173	10 Sep 2014	13:00
20	100	D	40,633	-122,297	30 Aug 2017	18:00

## 2.5. Синтез на експериментален математически модел за прогнозиране на вероятността от възникване на горски пожар. Резултати и анализи.

В тази точка са представени резултатите, получени при процеса на генериране на бинарния логистичен модел чрез софтуерната платформа SPSS. [70][90][99] В табл. 2.7. са показани изходните данни, описващи всички променливи, включени в модела със съответните им характеризиращи параметри.

Табл. 2.7. Променливи, включени в крайното уравнение на бинарната логистична регресия -“Variables in the equation”

Step 1 <sup>a</sup>	B	S.E	Df	Sig.	Exp(B)	95% C.I. for EXP(B)	
						Lower	Upper
Temp, [C]	-0,539	0,521	1	0,301	0,584	0,210	1,619
RH, [%]	0,046	0,172	1	0,791	1,047	0,747	1,466
SolR,[w/m2]	-0,003	0,003	1	0,189	0,997	0,991	1,002
Rain, [mm]	13,211	5,852	1	0,024	546451,084	5,702	>52*10 <sup>9</sup>
MinT, [C]	0,301	0,281	1	0,285	1,351	0,779	2,345
MaxT, [C]	1,159	0,604	1	0,055	3,187	0,975	10,414
Wind, [m/s]	-2,005	1,343	1	0,136	0,135	0,010	1,874
FM1, [%]	-2,767	1,289	1	0,032	0,063	0,005	0,786
FM10, [%]	0,811	0,761	1	0,286	2,251	0,507	10,001
FM100, [%]	2,107	0,770	1	0,006	8,222	1,818	37,185
FM1000,[%]	-2,039	0,704	1	0,004	0,130	0,033	0,517
KBDI	-0,015	0,007	1	0,022	0,985	0,973	0,998
Constant	-6,430	10,773	1	0,551	0,002		

Изходното уравнение на модела за бинарна логистична регресия е следното:

$$Y_i = B_0 + B_1 * X_{1i} + B_2 * X_{2i} + \dots + B_k * X_{ki}, \quad (2.29)$$

Където:

- $Y_i$  - Очаквана (прогнозираната) стойност на  $Y$
- $B_0$  – изчислената константа
- $B_1, B_2, \dots, B_k$  – изчислени теглови коефициенти на съответните променливите, включени в модела

- $X_{1i}, X_{2i}, \dots, X_{ki}$  - променливите, включени в модела (табл. 2.8.)

Получените стойности от табл. 2.7. за константата  $B_0$ , тегловите коефициенти  $B_k$  и променливите  $X_{ki}$ , заместяваме в уравнение (2.28) и получаваме прогнозираната стойност за изхода  $Y_i$  за всеки един от разглежданите случаи. За да получим от прогнозираната стойност на  $Y_i$ , вероятността (P), трябва да направим следните преобразувания:

Тъй като:

- P е вероятността дадено събитие Y да настъпи и приема стойност 1; ( $Y=1$ )
- $(1 - P)$  – е вероятността дадено събитие Y да не настъпи и приема стойност 0; ( $Y=0$ )

Следователно отношението на вероятността на едно събитие към вероятността на едно другото събитие, има следния вид:

$$\frac{p}{1-p} \quad (2.30)$$

Коего от своя страна е коефициента *odds*.

Следователно:

$$\log odds = \log \left[ \frac{p}{1-p} \right] = \text{logit}(p) = Y \quad (2.31)$$

Ако заместим Y със сумата от произведенията на променливите и съответните им теглови коефициенти, получени от бинарната регресия, ще получим:

$$\log \left[ \frac{p}{1-p} \right] = Y = B_0 + B_1 * X_{1i} + B_2 * X_{2i} + \dots + B_k * X_{ki} \quad (2.32)$$

След експониране на двете страни уравнението има следния вид:

$$\left[ \frac{p}{1-p} \right] = e^Y = e^{B_0 + B_1 * X_{1i} + B_2 * X_{2i} + \dots + B_k * X_{ki}} \quad (2.33)$$

Оттук извеждаме вероятността P в следния математически израз:

$$P = \frac{e^Y}{1+e^Y} = \frac{e^{B_0 + B_1 * X_{1i} + B_2 * X_{2i} + \dots + B_k * X_{ki}}}{1 + e^{B_0 + B_1 * X_{1i} + B_2 * X_{2i} + \dots + B_k * X_{ki}}} \quad (2.34)$$

Резултатите показват, че входните параметри с най-голяма значимост са типа на растителността и съдържанието на влага, описани от параметрите на „тежките“ горими органични материали (Heavy fuel moisture) - FM100 и FM1000. Съответно стойности на статистическата значимост Sig. при FM100 и FM1000 са равни на Sig.=0,006 и Sig.=0,004, което е значително по-малко от необходимият критерий, при който Sig. << 0,05. Следващите по значимост параметри са количеството дъждовни валежи (Rain), следвани от максималната дневна температура (maxT). Най-малко влияние върху разработения модел има относителната влажност на въздуха (RH).

След определяне на коефициентите и най-значимите параметри крайният израз за  $Y$  може да бъде записан, както следва:

$$Y = -6.43 + 13.211 * Rain + 1.159 * MaxT - 2.005 * Wind - 2.767 * FM1 + 2.107 * FM100 - 2.039 * FM1000 - 0.015 * KBDI$$

Вятърът (Wind), въпреки че има ниска статистическа значимост според данните в табл. 2.8., е включен в модела, тъй като има силна обратно пропорционална връзка с типа на горимите органични материали (FM100 и FM1000) и със съдържанието на влага в растителността. Също така параметърът „вятър“ е и основният фактор, влияещ на разпространението на пожарите.

## 2.6. Изводи

1. За изследване на взаимовръзката между независимите променливи, представляващи 12 метеорологични и топографски параметри (температура на въздуха, относителна влажност на въздуха, слънчева радиация, количество валежи, скорост на вятъра, минимална температура за деня, максимална температура за деня, както и пет индекса, описващи типа на горимите природни материали и водното съдържание в тях - FM1, FM10, F100, FM1000, KBDI), и зависимата променлива, отразяваща наличието на пожар, е избран статистическият нелинеен метод на бинарната логистична регресия.

2. Функционалната зависимост между зависимата величина и предикторите е определена по метода на максималното правдоподобие (Maximum Likelihood Estimation). Функционалната връзка е изразена чрез вероятността (Pf) за възникване на пожар.

3. Проверката на хипотезата за статистическа значимост на независимите променливи е направена чрез метода на хи-квадрат ( $\chi^2$ ). Вследствие на проверката са определени 7 от общо 12 независимите променливи като статистически значими за модела.

4. Установено е, че факторите с най-голяма значимост за възникване на пожар са типа на растителността и съдържанието на влага (FM100 и FM1000), докато най-слабо влияние имат относителната влажност на въздуха и вятъра.

5. Общата успеваемост на разработения регресионен модел за прогнозиране на вероятността от възникване на горски пожари показва висока степен на точност -  $g = 84,8$  %. Според модела броят на правилно определените прогнози възлиза на 76 от общо 91 изследвани случая.



6. Повишаване на точността може да се постигне чрез прилагане на изкуствено-невронен подход, който има предимството да обработва повече данни за обучение на мрежата като използва, както исторически записи, така и данни в реално време с цел коригиране на коефициентите на модела и повишаване на успеваемостта.

### **ГЛАВА III.**

## **РАЗРАБОТВАНЕ НА ИНТЕЛИГЕНТЕН ИЗЧИСЛИТЕЛЕН МОДЕЛ ЗА ОЦЕНКА И ПРОЗНОЗИРАНЕ НА ВЕРОЯТНОСТТА ОТ ВЪЗНИКВАНЕ НА ГОРСКИ ПОЖАРИ**

### **3.2. Синтез на експериментален интелигентен модел за прогнозиране на вероятността от възникване на горски пожар чрез използването на изкуствена невронна мрежа (ANN)**

За анализ на измерената информация чрез интелигентни модели за нуждите на прогнозиране на пожари е необходимо създаването на общ подход, който да гарантира висока степен на точност. Чрез него се генерира алгоритъм при ограничена извадка от входни данни, основана на голям брой измерени и експертно оценени променливи, използвайки проста по структура, бързосходима, многослойна невронна мрежа. Предимство на подхода е неговото приложение и в други приложни задачи. Обучената невронна мрежа за прогнозиране на вероятност от възникване на горски пожар е разработена въз основа на метеорологични и топографски особености за района на Уитмор, Калифорния.

За целта като най-подходящ предиктор за оценка на горската пожароопасност е избрана двуслойна невронна мрежа с 5 неврона в скрития слой и един неврон в изходния слой. Функцията на активиране на невроните в скрития слой е нелинейната диференцируема логаритмична сигмоидална функция (logsig). Активационната функция на единичния изходен неврон е функция от линеен тип. Обучението се извършва чрез алгоритъма за обратно разпространение на грешката - Backpropagation of error algorithm (BP).

#### **3.2.1. Събиране и предварителна обработка на данни за обучение на ANN**

Извадката от данни, използвана за генериране, обучение и валидиране на невронната мрежа за предсказване на горски пожар, описва 12 от най-големите горски пожари в района на Уитмор, Северна Калифорния, САЩ. Извадката е извлечена за

дългосрочен период от 10 години между 2007 г. и 2017 г. Входни данни, използвани за генерирането на модела, са: атмосферна температура (Temp), относителна влажност на въздуха (RH), слънчева радиация (SolR), количество дъждовни валежи (Rain), максималната температура през деня (maxT), средната скорост на вятъра в рамките на 10 минути преди измерване (Wind), 3 индекса, описващи типа на горимите природни материали и воднотосъдържание в него (FM1, F100, FM1000, KBDI), както и индекса на засушаване на Keetch Byram (KBDI).

Всеки  $i$ -ти ( $i=1\div N$ ) запис е комбинация от 10-те входни променливи, които съставят следния вектор  $pi_{10}$ :

$$pi_{10} = [Temp \ RH \ SolR \ Rain \ maxT \ Wind \ FM1, \ FM100, \ FM1000 \ KBDI].$$

Всички  $N$  комбинации изграждат матрицата на входните данни за невронната мрежа  $PN_{x10} = [pi \ x10]$ .

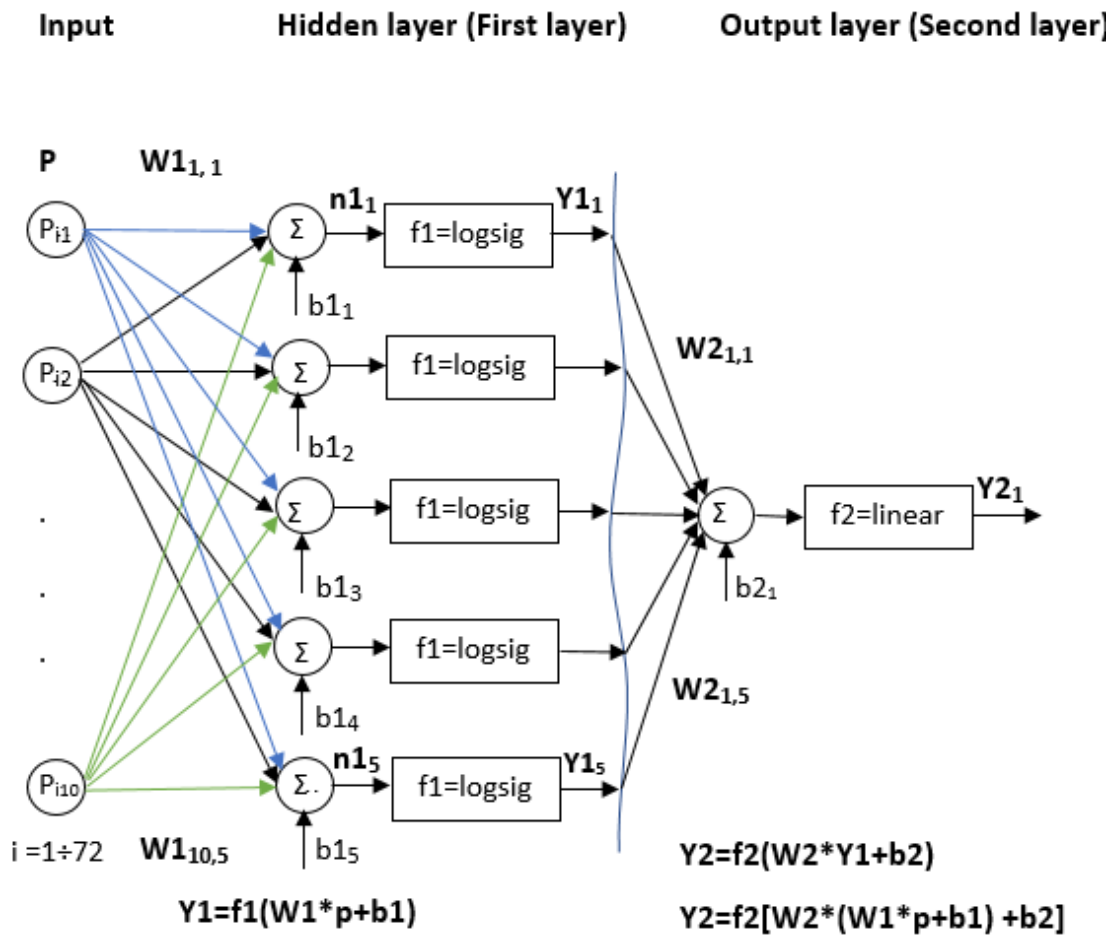
Входните данни се регистрират един път на всеки кръгъл час, 24 часа в денонощие.

Входните данни съдържат общо  $N_{total}=82$  вектора данни, от които  $pi_{10}$  – съответно 12 вектора, описващи реално събитие на пожар и 70 вектора, описващи събития с липса на пожар. Данните са избрани така, че да покриват различни условия и да гарантират, че няма линейна зависимост между различните вектори.

Събраните данни се разделят на  $N=72$  обучителна извадка в матрица  $P_{72 \times 10}$ , използвана за моделиране и обучение на невронната мрежа и  $N^v=10$  валидираща извадка в матрица  $P^v_{10 \times 10}$ , използвана за процеса на валидиране на мрежата. Обучителна и валидираща извадка, възлизат съответно на 87% и 13% от общия брой входни данни, използвани за изграждането на модела. Съответните целеви вектори  $T_{72 \times 1}$  и  $T^v_{10 \times 1}$  се състоят от изчислената за всеки вектор вероятност за пожар  $P_f$  [112]. Векторите  $P_f$  са получени с помощта на бинарен логистичен регресионен анализ за прогнозиране опасността от възникване на горски пожар, оценявайки възможните функционални зависимости между метеорологичните условия и специфичните характеристики на растителността за региона.

### **3.2.2. Обучение на невронната мрежа**

Обучението на избраната двуслойна праволинейна невронна мрежа за разработване на модел за прогнозиране на опасността от възникване на горски пожари, реализирано чрез алгоритъма BP, е направено с цел минимизиране на средната квадратична грешка (Mean Square Error - MSE) на функцията на загубата на изходния слой [113][114]. Блоквата схема на модела е представена на фиг.3.4.

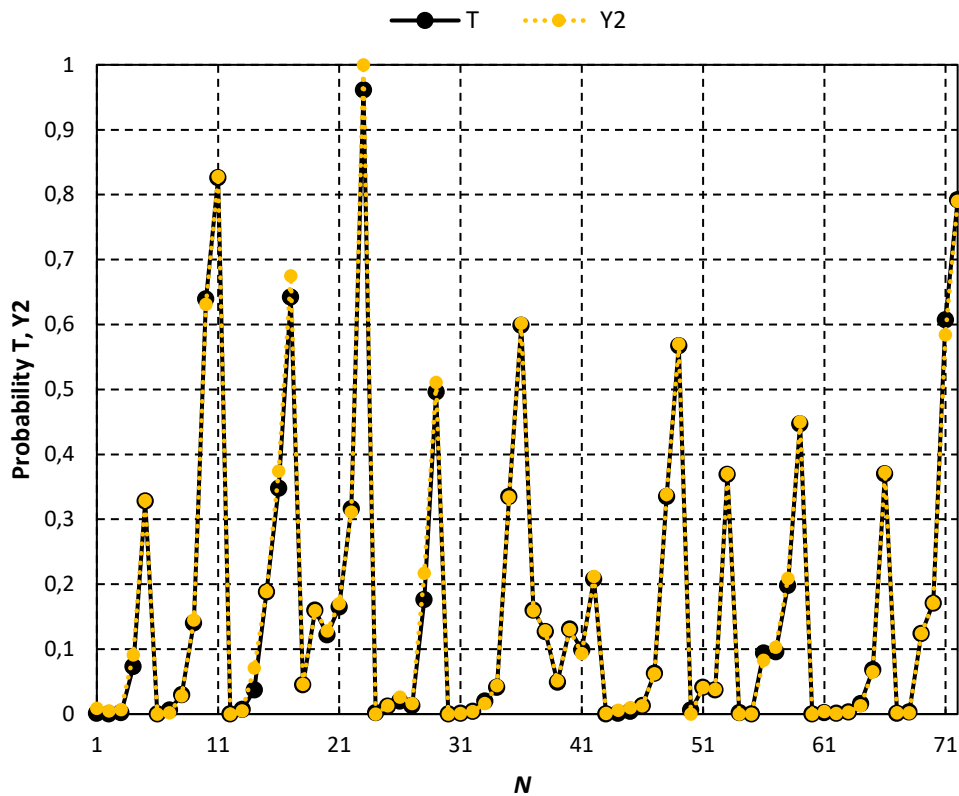


Фиг.3.4. Двуслойна невронна мрежа с 5 неврона в скрития слой и 1 неврон в изходния слой

За проектирането, инициализирането, обучението и валидирането на невронния модел е използван програмния продукт MATLAB™ [115].

Целевият изход  $T^{v_{72 \times 1}}$  и изчисленият изход от невронната мрежа са представени ( $Y2$ ) на фиг. 3.7. като функция от поредния номер  $Ni$ ,  $i=1 \div 72$  на съответния запис от входния набор данни.

Целевият изход ( $T$ ) е изобразен с плътна непрекъсната черна линия, докато изходът от невронната мрежа ( $Y2$ ) – с пунктирна жълта линия. На схемата може да се види, че изходът на невронната мрежа  $Y2$  следва плътно линията на целевият изход ( $T$ ) с висока степен на точност - по-малка от 5% разлика. Най-голяма разлика между двата изхода се наблюдава при записите  $Ni = 12$  и  $Ni = 23$ , където грешката остава по-малка от зададената стойност от 5%.



Фиг. 3.7. Целевият изход (Т) и изходът от невронната мрежа (Y2) при процеса на обучение

### 3.2.3. Валидиране на разработения модел на невронна мрежа

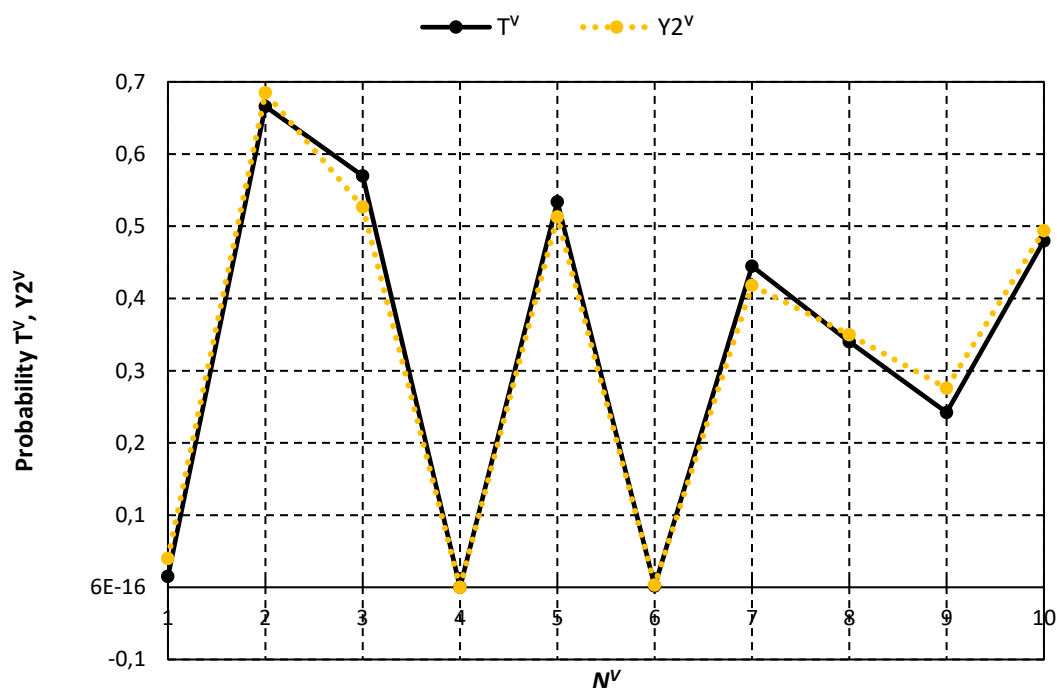
Валидирането на проектираната невронна мрежа за предсказване на горски пожар се осъществява чрез проверка на точността на прогнозата. Проверката се прави с нови, неизползвани при обучението на мрежата, входни данни.

Входните данни за валидиране са представени от двойките входно-изходни вектори ( $P^v_{10 \times 10}$ ,  $T^v_{10 \times 1}$ ). Те включват 10 вектора, 2 от които, описват реални събития на пожар ( $P_f \sim 1$ ), а останалите 8 - събития с липса на пожар ( $P_f \sim 0$ ).

На фиг. 3.8. с плътна черна линия са изобразени целевите изходни стойности  $T^v$  за вероятността  $P_f$ , докато изходът от невронната мрежа ( $Y2^v$ ) е представен с пунктирана жълта линия.

Резултатът показва, че е постигната точност от 95% при процеса на валидиране. Най-голямата отчетена разлика между двата изхода, е регистрирана при запис  $N^v=3$ , където абсолютната грешка остава по-малка от 5%. Точността на модела за предсказване на горски пожари илюстрира степента на обучение на невронната мрежата и осигурява

надеждна оценка при бъдещо прогнозиране с нови, неизползвани при обучението, входни данни.



Фиг.3.8. Целевият изход ( $T^v$ ) и изходът от невронната мрежа ( $Y2^v$ ) при процеса на валидиране

Обучението и последващото валидиране са неразделни процеси. Тази последователност се повтаря в цикъл, докато абсолютната разлика (грешка), между изчисления изход от невронната мрежа ( $Y2$ ,  $Y2^v$ ) и съответният му целеви изход ( $T$ ,  $T^v$ ), не стане по-малка от 0,05 (5%). Това условие е валидно както за процеса на обучение, така и за процеса на валидиране.

### 3.3. Изводи

1. Предложената обща методология за разработване и валидиране на многослоен ANN предиктор за прогнозиране на горски пожари може да се приложи при анализиране на опасността от пожар за различни географски региони с различни входни данни.

2. Проектираната невронна мрежа за прогнозиране на вероятността от възникване на горски пожари аналитично описва нелинейната връзка между голям брой значими променливи и вероятността от възникване на пожар.

3. Предложената изкуствена невронна мрежа отчита метеорологичните показатели и специфичните характеристики на растителността за дадения район, имащи най-голяма функционална зависимост и допринасящи най-много за възникването на горски пожар.

4. Предикторът се състои от двуслойна ANN с нелинейна активационна функция в скрития слой и линейна активационна функция в изходния слой. Скритият слой се състои от 5 неврона, а обучението се постига чрез алгоритъма за обратно разпространение на грешката. Мрежата е обучена с избрана ограничена извадка от входни данни от 82 входни вектора за 10-годишен период в района на Уитмор, Северна Калифорния.

5. ANN предиктора осигурява бърза обработка на данните, за което допринасят простата архитектура, бързо сходящият метод за обучение и малкото количество правилно предварително обработени входни данни.

6. Обучената невронна мрежа прогнозира възникването на горски пожари с по-висока степен на точност (грешка <5%) от модела за регресионно прогнозиране, разгледан в глава 2.

## **ГЛАВА IV.**

### **СРАВНИТЕЛЕН АНАЛИЗ И ОЦЕНКА НА МОДЕЛИ ЗА ПРОГНОЗИРАНЕ НА ВЕРОЯТНОСТТА ОТ ВЪЗНИКВАНЕ НА ГОРСКИ ПОЖАРИ**

#### **4.1. Въведение**

За оценка на ефективността при прогнозиране на вероятността от възникване на горски пожар се извършва сравнителен анализ между 5 различни методологии. Статистическият модел на бинарен регресионен анализ ( $P_f$ ), разработен в глава II, и интелигентният модел на изкуствена невронна мрежа (ANN), разработен в глава III, са сравнени с три съществуващи индекси от американската система за оценка на горската пожароопасност -  $BI'$ ,  $ERC'$  и  $SFDI'$ . Сравнението е направено като са използвани едни и същи входни данни за определен период и район (района на Уитмор).

#### **4.2. Нормализиране на индексите BI, ERC и SFDI**

Нормализирането на данните се състои в премоделиране на числови колони до стандартен мащаб. Операциите за преобразуване на работните данни са процедури по

предварителна обработка, които допринасят за успеха на процеса по извличане на необходимата работна извадка от големи по обем бази данни. Процесът на нормализиране е способ за привеждане на данните във вид, отговарящ на един общ мащаб или модел, което позволя интерпретацията, сравнението и анализа на данни от различни източници или измервания [116-118].

За нормализиране на данните е използван линейният метод, при който минималната и максималната стойност от данните се извличат и всяка стойност се привеждане в диапазона [0;1], чрез следния математически израз:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}, \quad (4.1)$$

където

- $\min(x)$  е минималната стойност в разглеждания диапазон от данни
- $\max(x)$  е максималната стойност в разглеждания диапазон от данни
- $x'$  е нормализираната стойност
- $x$  е съответната абсолютна стойност преди процеса на нормализиране

Чрез този метод се нормализират всички стойности на използваните параметри в диапазон [0;1].

Формула (4.2) може да се обобщи за всеки 2 произволни гранични стойности  $a$  и  $b$ . Математическият израз за обобщения вид на линейния метод за нормализация е, както следва:

$$x' = a + \frac{(x - \min(x)) * (b - a)}{\max(x) - \min(x)}, \quad (4.2)$$

където

- $a$  е началната стойност на ограничителния диапазон
- $b$  е крайната стойност на ограничителния диапазон

#### **4.3.1. Сравнителен анализ по критерий средна стойност и стандартно**

##### **отклонение**

Средната стойност на всеки индекс описва колко добре последният индикира вероятността от възпламеняване на пожар. Въпреки това тълкуването на средната стойност при двата класа „Fire” и „No Fire”, ще бъде различно по значение. При положителния клас „Fire”, по-високата средна стойност на индекса „Average“ дефинира по-точна прогноза за реално възникване на горски пожар. За разлика от отрицателния клас „No Fire“, при който прогнозата ще бъде с по – висока точност, когато средната стойност на индекса е по-ниска. По-ниските средни стойности за отрицателния клас

определят, че вероятността ( $P_f$ ) за възникване на пожар ще бъде по-ниска, което в случая е търсеният резултат.

Анализ на дескриптивна статистика на данните за 12-те събития с наличие на пожар (Fire), направени за всеки от петте индекса ( $ERC'$ ,  $BI'$ ,  $SFDI'$ ,  $P_f$  и ANN), е представен в табл. 4.2. Анализът е извършен чрез статистическия софтуерен продукт “Statgraphics Centurion XV” [119].

Може да се види, че сред всички 5 модела, двата индекса с най-високи средните стойности са  $BI'$  и  $ERC'$ , съответно  $BI'_{Average}=0,727$  и  $ERC'_{Average}=0,760$ , което ги определя като двата индекса с най-добра способност за прогнозиране на наличие на пожар.

Табл. 4.2. Дескриптивна статистика при клас “Fire” за индекси  $ERC'$ ,  $BI'$ ,  $SFDI'$ ,  $P_f$  and ANN

<b>FIRE</b>	<b><math>P_f</math></b>	<b><math>BI'</math></b>	<b><math>ERC'</math></b>	<b><math>SFDI'</math></b>	<b>ANN</b>
Count	12	12	12	12	12
Average	0.598	0.727	0.760	0.566	0.606
Standard deviation	0.188	0.098	0.113	0.259	0.198
Minimum	0.369	0.576	0.592	0.194	0.370
Maximum	0.961	0.852	0.963	1.000	1.000
Std. skewness	0.814	-0.247	0.090	0.107	0.836
Std. kurtosis	-0.359	-0.899	-0.516	-0.777	-0.243

По-ниско стандартно отклонение ( $Standard\ deviation \equiv Std.$ ) показва, че стойностите на данните клонят към средната стойност „Average“, докато по-високото стандартно отклонение показва, че стойностите са разпръснати в широк диапазон. Следователно ниската стойност на Std. е търсеният резултат, определящ кой индекс е по-добър класификатор.

Сравнявайки стандартното отклонение, можем да направим заключение, че  $BI'$  ( $Std=0,098$ ) и  $ERC'$  ( $Std=0,113$ ) показват най-добри резултати в сравнение с другите индекси, тъй като при тях са отчетени най-ниските стойности. Най-неефективния показател и по двата критерия (средна стойност и стандартното отклонение) е



индексът *SFDI'*, тъй като отчита най-ниска стойност за „Average“ (0,566) и най-висока стойност за „Std.“ (0,259) сред всички разгледани индекси.

В таблица 4.3. е представен анализ на дескриптивна статистика на данните за 70-те събития с липса на пожар (No Fire). Анализът е направен за всеки от петте индекса (*ERC'*, *BI'*, *SFDI'*,  $P_f$  и ANN).

В този случай, за прогнозиране на отрицателния клас (No fire), ниските средни стойности (Average) определят прогноза с по-висока точност. Анализирайки по критерий средна стойност, можем да направим заключение, че  $P_f$  (Average=0,115) и ANN (Average=0,113) показват най-добри резултати в сравнение с другите индекси, като се наблюдава леко предимство на нервната мрежа пред регресионния модел.

Табл. 4.3. Дескриптивна статистика при клас “No Fire” за индекси *ERC'*, *BI'*, *SFDI'*,  $P_f$  and ANN

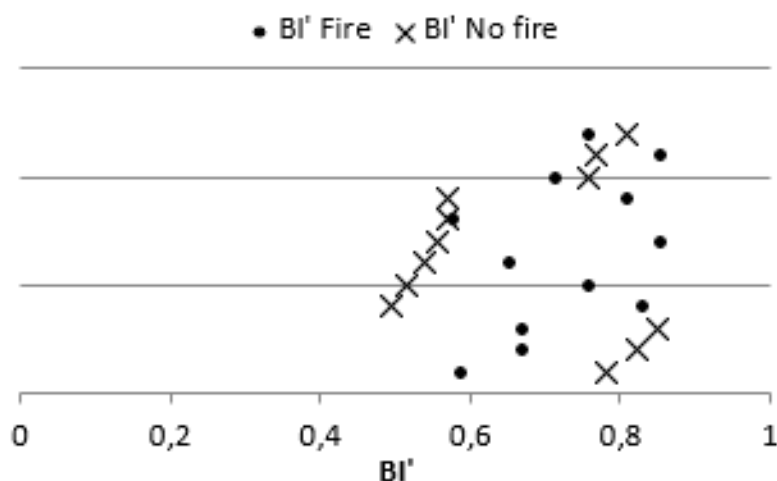
<b>NO FIRE</b>	<b><math>P_f</math></b>	<b><i>BI'</i></b>	<b><i>ERC'</i></b>	<b><i>SFDI'</i></b>	<b>ANN</b>
Count	70	70	70	70	70
Average	0.115	0.652	0.712	0.421	0.113
Standard deviation	0.162	0.109	0.120	0.258	0.157
Minimum	0.000	0.445	0.506	0.000	0.000
Maximum	0.639	0.848	0.951	0.979	0.631
Std. skewness	5.994	0.079	0.385	0.778	5.991
Std. kurtosis	4.248	-1.844	-1.478	-1.789	4.182

Резултатите показват също, че *BI'* и *ERC'* имат най-ниското стандартно отклонение, докато при *SFDI'* е отчетена най-високата стойност, както за положителния клас “Fire”, така и за отрицателния “No fire”. Анализът на разпределението на изчислените индекси показва, че и при двата класа (Fire, No fire), всички индекси са с нормалното разпределение на данните, с изключение на  $P_f$  и ANN за клас “No fire”. При нормално разпределение показателите Std. skewness и Std. kurtosis заемат стойности в интервала от -2 до +2.

Докато изискването за средната стойност „Average“ трябва да бъде напълно противоположно за двата класа „Fire” и „No Fire” (високи стойности за „Fire” и ниски за „No Fire”), то условието на което отговаря стандартното отклонение „Standard deviation“,

с цел постигане на по-точна прогноза, трябва да е едно и също в двата случая (ниски стойности за „Fire” и „No Fire”).

На фиг. 4.1. са илюстрирани някои от стойностите на BI при събития с наличие на пожар „Fire“ и при липса на пожар „No fire“. Резултатите показват, че близки стойности на индекса може да дефинират едновременно, както събития с липса на пожар „No fire“, така и такива, при които има реален пожар „Fire“.



Фиг. 4.1. Нормализирани стойности на индекса BI при състояние липса на пожар (No fire) и наличие на пожар (Fire)

Препокриване на стойностите за двата класа налага да дефинираме нов коефициент – коефициент на взаимноизключващи се състояния - Coefficient of mutually exclusive states determination (CMESD).

Коефициентът заема стойности в интервала [0;1] и се определя като разлика от средните стойности („Average“) на данните при клас “Fire” и клас “No fire”.

Математическият израз за коефициента на взаимноизключващи се състояния (CMESD) е както следва:

$$CMESD = Fire_{avr} - No\ fire_{avr} \quad (4.7)$$

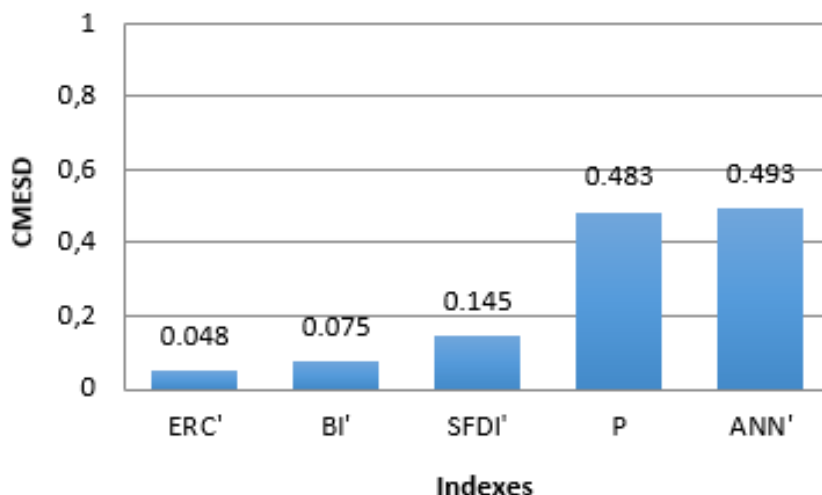
Нисък коефициент CMESD показва, че средните стойности за двата класа могат да са близки като показания. Следователно това е предпоставка двете категории да не могат ясно да се категоризират. Моделите за прогнозиране на горски пожари със ниски стойности на CMESD са най-неточните класификатори и при двата класа, в резултат на

което се затруднява целият оперативен процес по мониторинг, прогнозиране и предотвратяване на възникването на горски пожар.

Високите стойности на CMESD дефинират по-точна оценка за определяне на двата класа “Fire” и “No fire”. Генерирането на модели за прогнозиране на горски пожари с високи стойности на коефициента CMESD е целевият резултат, който осигурява висока точност на класификация на двата класа “Fire” и “No fire”.

Коефициентът на взаимноизключващи се състояния (CMESD) е изчислен и представен на фиг. 4.2. за всеки от петте индекса за прогнозиране на горски пожар BI', ERC', SFDI', P<sub>f</sub> и ANN.

Резултатите показват, че невронната мрежа (ANN) със стойност на коефициента CMESD=0,493 и регресионния анализ (P<sub>f</sub>) - със CMESD=0,483, са по-добри класификатори спрямо другите три индекса.



Фиг. 4.2. Стойности на коефициента на взаимноизключващи се състояния - Coefficient of mutually exclusive states determination (CMESD) за индексите BI', ERC', SFDI', P<sub>f</sub> и ANN.

#### 4.3.2. Сравнение по критерий средноквадратична грешка (RMSE)

Корен квадратен от средноквадратичната грешка (RMSE) се използва като мярка за точност на цифров индекс и е един от най-точните критерии за оценка на грешките на получените резултатите от измерванията. В прогнозирането, се използва за избор на модел, който дефинира по-малки прогнозни грешки. RMSE е корен квадратен на сумата от разликите между прогнозирана стойност и реално наблюдаваната за всички n измервания, повдигната на квадрат и разделена на размера на извадката.

Математическият израз за RMSE се описва със следната формула:

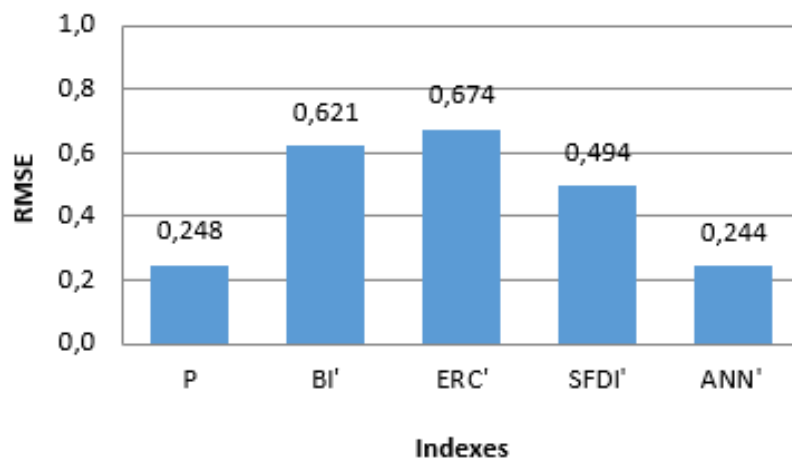
$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^n (Pi - Oi)^2}, \quad (4.8)$$

където

- $n$  е размерът на извадката (брой измервания)
- $Pi$  е прогнозирана стойност (predicted value) на  $i$ -тото измерване от извадката
- $Oi$  е реално наблюдаваната стойност (observed value) на  $i$ -тото измерване от извадката

Колкото по-голяма е грешка RMSE, толкова по-голяма е разликата между прогнозираните и наблюдаваните стойности, което означава, че разглежданият модел не се описва добре в извадката от данни, които са обект на анализ. Следователно колкото по-малка е RMSE, толкова по-добре моделът описва функцията на данните.

На фиг. 4.3. е илюстриран сравнен анализ по критерий средноквадратична грешка (RMSE) на петте модела за прогнозиране на вероятността от възникване на горски пожар след нормализиране на стойностите на индексите.



Фиг. 4.3. Средноквадратична грешка (RMSE) за петте индекса за прогнозиране на възникване на горски пожари BI', ERC', SFDI', P<sub>f</sub> и ANN.

Резултатите показват, че индексът ERC' има най-голяма средноквадратична грешка (RMSE=0,674), следван от BI' и SFDI', съответно с (RMSE=0,621) и (RMSE=0,494).

Модели за прогнозиране на горски пожари с най-ниски стойности на средноквадратичната грешка са регресионния модел P<sub>f</sub> (RMSE=0,248) и невронната мрежа ANN (RMSE=0,242).

#### **4.4. Изводи**

1. При клас „Fire“, определящ събитията с наличие на пожар, нормализираните индекси  $BI'$ ,  $ERC'$ ,  $SFDI'$ , а също така бинарната регресия  $P_f$  и неврония модел ANN следват модела на нормално разпределение и резултатите от тях могат да бъдат представени със стандартна неопределеност. За клас „No Fire“, дефиниращ събитията с липса на пожар, индексите  $BI'$ ,  $ERC'$ ,  $SFDI'$ , също могат да бъдат представени със стандартна неопределеност.

2. Резултатите от сравнителният анализ показват, че моделите с индекси  $BI'$  и  $ERC'$  са добри класификатори по отношение на клас „Fire“, но представят незадоволителен резултат при клас „No fire“. За разлика от тях, моделът с бинарната регресия ( $P_f$ ) и невронната мрежа (ANN) показва значително по-добри резултати при прогнозиране на събитията с липса на пожар „No fire“.

3. Въз основа на дефинираният коефициент на взаимноизключващи се състояния CMESD, бинарният регресионен модел ( $P_f$ ) и невронната мрежа (ANN) демонстрират най-добрите резултати за класификация и при двете класа - „Fire“ и „No fire“.

4. Разработените модели  $P_f$  и ANN представят значително по-добри резултати по критерий средноквадратична грешка (RMSE) в сравнение с другите 3 индекса.

5. От представените резултати може да се заключи, че двата модела, разработени и изложени в настоящето изследване (регресионният модел  $P_f$  и невронната мрежа ANN) могат да бъдат възприети като водещи за прогнозиране и ранно откриване на горски пожари.

#### **ПРИНОСИ В ДИСЕРТАЦИОННИЯ ТРУД (НАУЧНО – ПРИЛОЖНИ)**

1. Синтезиран е бинарен регресионен модел за предсказване на горски пожари за района на Уитмор, Северна Калифорния. Статистически е оценено влиянието на всяка от независимите променливите участващи в модела (количество валежи, максимална температура за деня, скорост на вятъра и четирите индекса, описващи типа на горимите природни материали и водното съдържание в тях -  $FM1$ ,  $F100$ ,  $FM1000$ ,  $KBDI$ ) върху вероятността за възникване на пожар. Установено е, че за изследвания регион, типът на растителността и съдържанието на влага в нея са с най-голяма статистическа значимост, следвани от количеството валежи. Статистически е оценена точността на синтезирания модел, като е установено, че общата успеваемост е 84,8%.

2. Представен е обобщен алгоритъм за разработване и валидиране на експериментален интелигентен модел за прогнозиране на вероятността от възникване на горски пожар чрез използването на многослойна изкуствена невронна мрежа (ANN). Синтезиран, на базата на предложения алгоритъм е модел на предиктор, основан на голям брой реално измерени и експертно оценени променливи, които описват различни характеристики на региона. Обучената невронна мрежа използва представителна, но ограничена по обем извадка от входни данни, състояща се от 82 линейно независими входни вектора и прогнозира възникването на събития с по-висока точност спрямо модела за регресионно прогнозиране.

3. Извършен е сравнителен анализ между обучената невронна мрежа, бинарния регресионен модел и нормализираните общоприети индекси за предсказване на пожари (Energy Release Component (ERC), Burning Index (BI) и Severe Fire Danger Index (SFDI)), базиран на реално измерени стойности от едни и същи станции в един и същи период от време, като е установено разсейване на резултатите при всеки един от сравняваните методи.

4. Анализирани са статистически оценки за центъра на групиране и на дисперсията на получените резултати за всеки от изследваните методи и е установено, че бинарният регресионен модел и невронната мрежа превъзхождат широко разпространените индекси при прогнозиране на отрицателен клас събития (липса на пожар). Оценена е и средноквадратична грешка (RMSE) между изчислените и измерените резултати за всеки от петте изследвани метода, като е установено, че бинарният регресионен модел и невронната мрежа са по-точни методи спрямо съществуващите индекси.

5. Дефиниран е коефициент на взаимоизключващи се състояния (CMESD), като е установено, че бинарният регресионен модел и невронната мрежа са по-точни класификатори спрямо съществуващите индекси.

6. Установено е, че резултатите генерирани от бинарния регресионен модел и обучената невронна мрежа и при двата класа събития (наличие и липса на пожар), могат да се представят със стандартна неопределеност, като това твърдение не може да се приеме за валидно при съществуващите индекси за прогнозиране.

## **СПИСЪК НА НАУЧНИТЕ ПУБЛИКАЦИИ, СВЪРЗАНИ С ДИСЕРТАЦИОННИЯ ТРУД**

- [1] Nikova H. „Forest Fire Warning Systems – Types and Principles“, Proceedings of XIX International conference on „Challenges in Higher Education and research in the 21th century“, 2023, vol.19, pp. 63-68, ISSN: 2683-0337
- [2] Nikova H., R. Deliyski. “Binary Regression Model for Automated Wildfire Early Prediction and Prevention”, 11th International Scientific Conference on Computer Science (COMSCI), Sozopol, Bulgaria, 2023, DOI: 10.1109/COMSCI59259.2023.10315856 (SCOPUS)
- [3] Nikova H., S. Yordanova, R. Deliyski. „Application of Artificial Neural Network in Wildfire Early Prediction Systems“, WSEAS Transactions on Environment and Development, vol. 19, pp. 1410-1420, 2023, ISSN: 1790-5079, E-ISSN: 2224-3496, DOI: [10.37394/232015.2023.19.128](https://doi.org/10.37394/232015.2023.19.128) (SCOPUS)
- [4] Nikova H., R. Deliyski, T. Tashev. „Analysis of Wildfire Early Prediction Models“, Proceedings of XIX International conference on „Challenges in Higher Education and research in the 21th century“ 2023, vol.19, pp. 78-82, ISSN: 2683-0337

## **SUMMARY**

The present thesis has been developed under supervision of the Technical University of Sofia, Bulgaria and is focused on the study of forest fire prediction occurrence, monitoring, early detection.

The main objective pursued is to gain understanding about the development and research of models for forecasting the probability of forest fires based on information from a real measurement system. For this purpose, the following 4 investigational tasks are performed:

Development of an experimental statistical model for predicting the probability of forest fires occurrence and determining the influence of the main variables affecting the probability; Analysis and assessment of the adequacy of the developed statistical model; Development and validation of an experimental intelligent model for forest fire probability prediction using a multi-layer artificial neural network (ANN); A comparative analysis between the developed models and generally accepted fire danger indexes is performed.

A statistical model based on binary regression analysis and an intelligent model based on artificial neural network approach are generated for estimating the probability of forest fire occurrence. The input data cover 10 years (from 2007 to 2017) and contain meteorological and topographical parameters including current temperature (Temp), relative humidity (RH), total solar radiation for the day (SolR), rain in terms of total precipitation amount (Rain), the maximum temperature during the day (maxT), the minimum temperature during the day (minT), the average wind speed within 10 min period before measurement (Wind), fuel

moisture - FM1, FM10, FM100, FM1000 as well as Keetch Byram Drought Index (KBDI). The input data is registered in the region of Whitmore, North California, USA.

Programming code implementations have been developed in MATLAB environment for the artificial neural network approach generated in this study. While to build the binary regression model the IBM statistical software platform SPSS is used.

The intelligent approach consists of a two-layer ANN with nonlinear activation functions of the hidden neurons and a linear activation function of the output neuron trained by backpropagation of error learning algorithm. The ANN is accepted to have 5 neurons in the hidden layer and one neuron in the output layer. The nonlinear activation function of the neurons in the hidden layer is differentiable logsig with output in the range [0, 1]. The comprehensive and low-complexity algorithm of the network ensures fast processing. To this, contribute the ANN simple architecture, the fast-converging training method, and the small amount of properly pre-processed training data.

The overall success rate of the developed regression model for predicting the probability of forest fire occurrence shows a high degree of accuracy -  $g=84.8\%$ . The trained neural network predicts forest fire occurrence with a higher degree of accuracy (error  $<5\%$ ) than the generated binary regression model.

A comparative analysis has been performed between the trained neural network, the generated binary regression model and three generally accepted mass-used indexes for forest fire prediction (ERC, BI and SFDI) based on real measured data from the same region and period. The root mean square error (RMSE) between the calculated and measured results for each of the five investigated methods has been also evaluated, and the binary regression model and the neural network were found to be more accurate methods compared to the existing indexes. The results of the analysis considering the "Standard deviation" and "Average" of a dataset show that the indexes BI' and ERC' are good classifiers in terms of class "Fire", but present an unsatisfactory result in class "No fire". In contrast, the model with the binary regression ( $P_f$ ) and the neural network (ANN) shows significantly better results in predicting the events with no fire "No fire". From the presented results it can be concluded that the two models developed and exposed in the present study (the binary regression model and the ANN) can be perceived as leading for forest fire prediction and early detection.