



**БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ  
ИНСТИТУТ ПО ИНФОРМАЦИОННИ И КОМУНИКАЦИОННИ  
ТЕХНОЛОГИИ**

**Десислава Николова Бояджиева**

**Комбиниран подход за разпознаване на он-лайн подписи**

**А В Т О Р Е Ф Е Р А Т**

**НА ДИСЕРТАЦИЯ**

**за присъждане на образователната и научната степен „доктор“**

**по научна специалност 01.01.12 „Информатика“**

**Научен ръководител: доц. д-р Георги Глухчев**

**София, 2014 г.**

Дисертацията е обсъдена и допусната до защита на разширено заседание на секция „Обработка на сигнали и разпознаване на образи“ на ИИКТ-БАН, състояло се на 17.01.2014 г.

Дисертацията съдържа 129 стр., в които 18 фигури, 19 таблици и 7 стр. литература, включваща 108 заглавия.

Защитата на дисертацията ще се състои на .....2014 г. от ..... часа в зала ... на блок ... на ИИКТ-БАН на открито заседание на научно жури в състав:

1. ....
2. ....
3. ....
4. ....
5. ....

Резервни членове:

1. ....
2. ....

Материалите за защитата са на разположение на интересуващите се в ИИКТ-БАН, ул. „Акад. Г. Бончев“, бл. 25А, стая 215.

Автор: маг. *Десислава Николова Бояджиева*

Заглавие: **КОМБИНИРАН ПОДХОД ЗА РАЗПОЗНАВАНЕ НА ОН-ЛАЙН ПОДПИСИ**

# Обща характеристика на дисертацията

## Актуалност на темата

Верификацията и идентификацията по подпис, обединени от общия термин разпознаване по подпис, са задачи от областта на разпознаването на образи и анализа на документи. Целта на верификацията е потвърждаването или отхвърлянето на заявена самоличност, а на идентификацията - установяването на самоличността на даден потребител. В зависимост от начина на събиране на подписи, системите за разпознаване по подпис са статични (оф-лайн) и динамични (он-лайн). При първите подписът се привежда в цифров вид чрез сканиране или заснемане с фотокамера след като вече е бил положен върху хартия и по-нататък се работи с неговото полутоново статично изображение, докато при вторите подписите се полагат директно чрез дигитална писалка върху устройства като графичен таблет или друг вид сензорен екран, като по този начин автоматично се получава едновременно и статична, и динамична информация за подписа.

Тъй като подписът е често употребявано, удобно, сигурно и широко възприето средство за установяване на самоличност, с неинвазивно, бързо, лесно и естествено за потребителите въвеждане в биометричните системи, то през последните години се наблюдава засилен интерес към разработките в областта. Разработват се нови методи и алгоритми, част от които се внедряват в практически приложения [7, 13, 33, 40, 41, 42], като се наблюдава превес на он-лайн пред оф-лайн методите за разпознаване по подпис. С изобретяването на все повече и повече нови устройства с дигитални писалки и сензорни екрани, които предоставят възможност още с полагане на подписа да се разполага с неговия дигитален формат, се откриват нови хоризонти пред областите на приложение на он-лайн системи за разпознаване по подпис. Така практически всяко приложение за достъп до система, ресурс или устройство, за което се изисква парола за достъп, може да се замени с он-лайн верификация по подпис, а он-лайн идентификацията по подпис успешно да се прилага в случаи на хартиен носител. Изобщо, широките области на приложение в създаването и внедряването на системи за установяване на самоличност определят актуалността на задачата в световен мащаб, а оттам и актуалността на дисертационната работа.

## Цели и задачи на дисертацията

Основната цел на дисертационния труд е ***Разработване на нов комбиниран метод за разпознаване на потребители по он-лайн подписи и реализирането му като софтуерна система.*** За изпълнение на поставената цел е необходимо да се решат следните задачи:

- Организиране на база данни от подписи;
- Извличане на известни и нови признаци;
- Избор на оптимални подмножества от признаци;
- Изследване на различни модели на невронни мрежи за верификация като елементи на комбиниран класификатор;
- Провеждане на експерименти за оценка на точността на класификация и сравнение с известни класификатори с използване както на собствена база данни за подписи, така и с публично достъпна такава;
- Разработка на софтуерно приложение за разпознаване на подписи с основни

функции: събиране на собствени и подправени подписи, визуализация на подписите, извличане на признаци, избор на признаци и верификация по тях.

Особеностите на задачата налагат определени ограничения, съобразяването с които прави някои от стандартните подходи не особено подходящи. Тези особености се отнасят преди всичко към количеството на експерименталните данни, използвани за обучение и тестване на класификатора и характера на използваните признаци. За разлика от други задачи тук не е възможно събирането на голям брой образци от подписи на евентуалните потребители. Изискването за многократно подписване е неприятно. Броят на желаещите намалява значително и ако се иска неколкостранно повторение на сесиите. Добавяйки и нежеланието на повечето хора да предоставят подписите си за експерименти, страхувайки се от злоупотреба, броят на събраните подписи се свежда до няколко десетки. Сравнително по-големи бази от подписи могат да бъдат събрани в университетите. Още по-голямо затруднение представлява събирането на умели фалшификати, необходими за повишаване на точността на класификатора.

Изборът на признаци също не е тривиален по две причини. Първата е свързана с малкия брой опитни данни. От теорията на разпознаването на образи е известно, че при фиксиран брой опитни данни увеличаването на броя на признаците не само не води до повишаване на точността на разпознаването, но я и намалява. Това означава, че ако при малък брой данни имаме голям брой признаци, те трябва да се намалят, т.е. да се избере подмножество, което да бъде достатъчно информативно. Намаляването на броя на признаците е свързано и с намаляване на изчислителното време. Втората причина при избора им е необходимостта от разбираем физически смисъл. Това се налага при съдебни експертизи, където резултатът от машината не се приема от съда за доказателство, ако зад него не застане експерт. (Обяснението е, че от машината не може да бъде търсена отговорност).

Независимо от тези сериозни практически ограничения, към всеки класификатор се предявяват изисквания за висока точност на разпознаване, изразяваща се в малък процент на двата вида грешки: дял на лъжливо положителните образци (FAR) и дял на лъжливо отрицателните образци (FRR). Допълнителна особеност в това отношение е различната цена на двете грешки, която трябва да бъде отчетена от класификатора. Много по-сериозна е първата грешка, която може да доведе до тежки последици за институцията, докато при втората ще бъде необходимо само полагането на допълнителен подпис.

Заедно с формулираните цел и задачи на дисертацията ще бъдат потърсени отговори и на следните въпроси.

- 1) Съществува ли конкретно, специфично подмножество от признаци за всеки потребител, отразяващо особеностите на неговия стил на подписване?
- 2) Има ли съществена разлика между резултатите, получени при верификацията с общо множество от признаци и индивидуално подмножество за всеки участник?
- 3) Ще се понижи ли точността на верификация с използването на умели фалшификати в качеството на отрицателни примери?

## **Методология на изследването**

Методологията на настоящите изследвания се основава на методи и подходи от следните области:

- Статистика – корелационен анализ, избор на регресионни променливи, крос валидация;

- Информатика - невронни мрежи, метод на  $k$ -най-близки съседи;
- Разпознаване на образи – схеми за верификация.

В дисертационния труд за тестване на предложеният комбиниран метод за разпознаване на он-лайн подписи са използвани образци на подписи от собствена база данни и от публично достъпна база от данни.

## Апробация на резултатите

Част от резултатите са докладвани на 4 международни конференции, 2 от които в чужбина (BioID\_MultiComm'09, PReMI'11) и 2 в България (UNITECH'10, 14th Int. Workshop on Generalized Nets). 1 статия е докладвана на научна конференция в България и една е приета за печат в списанието *Cybernetics and Information Technologies*.

## Списък на публикациите по дисертацията

1. Boyadzhieva D., Gluhchev G., Feature Set Selection for On-line Signatures using Selection of Regression Variables, In: Proceedings of the 4th International Conference on Pattern Recognition and Machine Intelligence PReMI'11, June 27-July 01 2011, Moscow, Russia, LNCS 6744, Springer-Verlag Berlin Heidelberg, 2011, ISSN 0302-9743, pp. 440-445, [http://link.springer.com/chapter/10.1007/978-3-642-21786-9\\_71#page-1](http://link.springer.com/chapter/10.1007/978-3-642-21786-9_71#page-1).
2. Boyadzhieva D., Gluhchev G., An approach to feature selection for on-line signatures, In: Proceedings of the 10th Anniversary International Scientific Conference UNITECH'10 Vol. I, Nov. 19-20, 2010, Gabrovo, Bulgaria, ISSN 1313-230X, pp. I-457...I-460.
3. Бояджиева Д., Извличане на признаци на подпис от графичен таблет, В: Научни трудове на Русенски университет "Ангел Кънчев", 2010, том 49, серия 3.2, Русе, България, ISBN 1311-3321, стр. 101-105, <http://conf.uni-ruse.bg/bg/docs/cp10/3.2/3.2-18.pdf>.
4. Dimitrova D., Gluhchev G., Pressure Evaluation in On-Line and Off-Line Signatures, In: Proceedings of the Joint COST 2101 & 2102 International Conference on Biometric ID Management and Multimodal Communication (BioID\_MultiComm'09), Sep. 16-18 2009, Madrid, Spain, LNCS 5707, Springer -Verlag Berlin Heidelberg, 2009, ISSN 0302-9743, pp. 207-211, [http://link.springer.com/chapter/10.1007/978-3-642-04391-8\\_27#page-1](http://link.springer.com/chapter/10.1007/978-3-642-04391-8_27#page-1).
5. Desislava Boyadzhieva and Georgi Gluhchev., *A GN model for on-line signature verification*, In: Proceedings of the 14th Int. Workshop on Generalized Nets, Burgas, Bulgaria, 29-30 November 2013, ISSN 1313-6860, pp. 65-70, <http://www.ifigenia.org/w/images/f/f6/IWGN2013-14-65-70.pdf>
6. Boyadzhieva D., Gluhchev G., *A Combined Method for On-Line Signature Verification* (приета за печат в списанието *Cybernetics and Information Technologies*, кн. 2, 2014 г).

## Съдържание на дисертацията

Настоящата дисертация се състои от списък на фигурите, списък на таблиците, речник на термините, увод, 5 глави, заключение, приноси, публикации по темата и библиографска справка. Първа глава е озаглавена "Разпознаване на подписи", втора глава - "Избор на признаци за он-лайн подписи", трета глава - "Класификация", четвърта – "Експерименти и резултати" и пета – "Софтуерно приложение". Основното съдържание е изложено на 129 страници. Включени са 42 формули, 18 фигури, 19 таблици и 7 стр. библиографска справка, включваща 108 заглавия.

## Глава 1. Разпознаване на подписи

Първа глава включва следните раздели.

### 1.1 Постановка на задачата за разпознаване на подписи

#### 1.1.1 Верификация и идентификация на самоличност

#### 1.1.2 Биометрични данни и подписът като биометрични данни

#### 1.1.3 Системи за разпознаване по подпис (СРП)

Разпознаването по подпис представлява процеса по потвърждаване на самоличността на потребител въз основа на неговия подпис като поведенческа биометрична характеристика [33] и е област на изследване в областта на разпознаването на образи и анализа на документи. В системите за разпознаване по подпис се фокусира върху извличането на характерните и специфични характеристики от образците на подписите за всеки потребител, които го разграничават възможно най-много от подписите на останалите потребители. Подписът като биометрична модалност по-често намира приложение в системи за достъп до определен ресурс и в системи за идентификация на базата на хартиен носител.

#### 1.1.4 Он-лайн и оф-лайн СРП

#### 1.1.5 Приложение на он-лайн СРП

#### 1.1.6 Видове подправени подписи в он-лайн СРП

За разлика от някои биометрични модалности, подписът може да бъде подправен с относителна лекота. Разграничават се следните типове подправени подписи: неумели (*random*), умели (*skilled*) и прости (*simple*). При първите фалшификаторът използва собствения си подпис вместо този, който ще подправи, докато при вторите фалшификаторът е запознат с собствения подпис и динамиката на изписването му. При простите фалшификати пък, имитаторът изхождайки от името на потребителя, чийто подпис ще подправи, създава подправения подпис. На Фигура 1.1 са представени собственият и трите типа подправени подписи за потребител.



Фигура 1.1 Собствен и три типа подправени подписа (неумел, умел и прост) за потребител

Степента на сходство на подправените подписи с собствените подписи зависи от протокола за събиране на подписи, от мотивацията на фалшификаторите, от техните умения да подправят подписи, от това доколко са запознати с подписите и по какъв начин се упражняват, както и от времето, с което разполагат за трениране. Събирането на подправени подписи от хора, които нямат опит във фалшифицирането на подписи, както се случва на практика, е причина за отсъствието на качествени фалшификати. Фалшифицирането на динамиката при подписване е значително по-трудно отколкото имитирането единствено на формата на подписа, тъй като на практика фалшификаторът рядко има достъп до динамичните характеристики на подписа.

### 1.1.7 Етапи в създаването на он-лайн СРП

Създаването на СРП (он-лайн или оф-лайн) преминава през следните етапи: събиране на подписи, предварителна обработка, извличане на признаци и избор на най-информативните сред тях, разработване на класифициращи правила и оценяване на вероятността за грешка при класификацията [1].

- 1) *Събиране на подписи* – Събирането на он-лайн подписи се осъществява чрез устройства, например графични таблети, PDA, Pocket PC, Tablet PC, умни телефони и др. Най-често с тях се измерват следните характеристики: а)  $x$ ,  $y$  координатите на върха на писалката; б) приложението от писалката натиск; в) времеви отпечатък; г) ъгъл на азимут на писалката (ротация на писалката по часовниковата стрелка около оста  $x$ , със стойности между  $0^\circ$  и  $359^\circ$ ); д) ъгъл на наклон на писалката по отношение на повърхността на графичния таблет (със стойности между  $0^\circ$  и  $90^\circ$ ).
- 2) *Предварителна обработка* – това е обработката, която се извършва, за да се подготвят подписите за следващия етап. Тя обикновено включва филтриране, за премахване на точките, не принадлежащи на подписа, трансляция на подписа в дадена точка, ротация, целяща подравняването на подписа хоризонтално, а понякога и нормиране, за да се стандартизира подписа по отношение на размера.
- 3) *Извличане на признаци* – Те се пресмятат като функции от извлечените характеристики. Признаците би трябвало да позволят разграничаването на истинските от подправените подписи и да са подходящи за автоматично сравнение. Намирането на добро дискриминационно множество от признаци предопределя до голяма степен успеха на всяка система за разпознаване.
- 4) *Класификация* - Основната задача в областта на разпознаването на образи е построяването на класификатор. При нея се извършва причисляване на входните образци към един от множество от  $N$  класа.
- 5) *Оценка на вероятността за грешка* - Качеството на биометричните системи се оценява в термините на оценки на грешката при вземане на решение (*decision error rates*). В качеството на оценки се използват FAR (*false accept rate*, дял на лъжливо положителните образци) и FRR (*false reject rate*, дял на лъжливо отрицателните образци).

### 1.1.8 Извличане на признаци и подходи за разпознаване по подписи

Признаците на подписи представляват функции от извлечените характеристики, а изборът на оптимално признаково множество е важна стъпка в създаването на СРП. От първостепенно значение е извлечените признаци да имат невисоки изисквания към изчислителни ресурси и да водят до добро разпознаване с ниска стойност на FRR и FAR. Различните СРП се различават по използваните признаци и метода на разпознаване. Повече от 100 признаци са използвани в разработките в областта [2]. В зависимост от критерия, по който се групират, признаците за подписи се делят на [5]:

- А) Според начина на събиране на подписите: статични и динамични;
- Б) Според обхвата: глобални и локални;
- В) Според начина на представяне: функционални и параметрични (скаларни).

### 1.1.9 Трудности при създаването на СРП

Подписът има някои характерни особености, които създават известни затруднения при създаването на надеждна СРП. На първо място следва да се отбележи, че представителите на подписите от един и същи потребител са подобни, почти еднакви, но не и идентични (съществува голяма вътре класова вариация), както и че

последователно положени подписи от един и същ потребител се различават по мащаб и по ориентация. Върху качеството на подписите оказват влияние и различните условия, при които те се полагат. Например набързо положения подпис в изправена поза би се различавал от подписа, положен докато участникът седи. Значение би оказала и употребата на необичайна писалка, с която потребителят не е свикнал, както и моментното психологическо състояние на участника. Трудностите при разпознаването произтичат от късата дължина на някои подписи, която би могла да разкрие малка информация, а оттам да доведе до неточното им разпознаване. Също така, потребители с подобни или еднакви имена би могло да имат сходни подписи по геометрична форма [5].

Трудности произтичат и от процеса по набавянето на собствени и подправени подписи. Обикновено СРП разполага с ограничен брой подписи. Създаването на тестова база от данни от подписи, която да е представителна за реални приложения е тежка задача, поради трудностите, които обичайно се срещат при набирането на доброволци, които да се съгласят да представят известен брой подписи и то в различни сесии. Повечето от хората не биха се съгласили подписите им да се съхраняват в бази от данни и да се предоставят за фалшифициране.

## 1.2 Критичен обзор

В тази част е извършен преглед на разработките в областта на разпознаването на подписи от последните години. Обърнато е внимание на обзорните статии по темата и са разгледани различни подходи за решаване на задачата, свързани с настоящата дисертация. Тези методи са дискутирани и са представени конкретни системи за разпознаване по подпис, проектирани върху тях. За всяка от системите са посочени използваните видове признаци и метода за разпознаване, вида на използваните фалшифицирани подписи, базата от данни, върху която са проведени експериментите и получената оценка на системата.

През последните 30 години са разработени голям брой методи и модели за верификация на он-лайн подписи, които се основават най-общо на следните методи: статистически подходи, динамично времево изкривяване (*Dynamic Time Warping*), скрит Марковски модел (*Hidden Markov Model*), невронни мрежи, комбинирани подходи [8] и класификатор по  $k$ -най-близки съседи.

На база на направения литературен обзор може да се направи извода, че текущите изследвания и разработки в областта на разпознаването на подписи са насочени основно към създаването на комбинирани подходи и по-добър избор на признаци, които да разграничават надеждно собствените от фалшифицираните подписи.

## 1.3 Цели на дисертационната работа

### 1.4 Резултати в Глава 1

1. Представена е биометричната модалност „подпис“, начините за получаване и възможните приложения.
2. Направена е характеристика на използваните признаци, етапите на изграждане на СРП и съпътстващите ги трудности.
3. Представени са основните методи за разпознаване: статистически, динамично времево изкривяване, скрити Марковски вериги, невронни мрежи, комбинирани подходи. Посочени са съответните литературни източници и данни за ефективността на методите.
4. Формулирана е целта на дисертацията и съпътстващите изпълнението ѝ задачи.



## Глава 2. Избор на признаци за он-лайн подписи

Втора глава се състои от следните раздели.

### 2.1 Предварителна обработка

Една от основните задачи в разпознаването на образи е изборът на признаци, по които да се извършва разпознаването. Те се избират от изследователя в зависимост от неговия опит, а също и от спецификата на разпознаваните обекти. Предварително не е известно кои са признаците, които биха довели до задоволително разпознаване и затова обикновено изследователят разглежда възможно най-голям брой признаци, стремейки се по този начин да не пропусне някой важен за разпознаването признак. Същевременно, ефективността на системата за разпознаване и изискванията към изчислителните ресурси зависят в голяма степен от броя на признаците [34]. Затова колкото е по-малък той, толкова е по-голяма скоростта на верификация [6].

Така възниква и задачата за избор на най-информативните признаци, решаването на която се състои в минимизиране на признаковото пространство, така че да се премахнат признаците, които са носители на излишна или повтаряща се информация.

Бихме могли да формулираме следните критерии към признаците, които да участват в признаковото множество за он-лайн подписи:

1. Да бъдат извлечени и пресмятани бързо и лесно;
2. Ако по някакъв начин станат достъпни от неоторизирани лица, да не позволяват лесното пресъздаване на подписа чрез тях;
3. Броят им да бъде достатъчно малък, за да не се забавят по-нататъшните изчисления и достатъчно голям, за да осигури надеждна верификация;
4. Желателно е да имат разбираем физически смисъл, което ще позволи установяването на фалшиви подписи от експерт в случай на съдебна експертиза.

Наред с избора на модел и вида на класификатора избраните признаци играят много важна роля за постигането на успех на всяка СРП.

#### 2.1.1 Предварителна обработка на данните от подпис

Както е известно, според вида на използвания хардуер (графичен таблет, PDA и др.) за събиране на подписи, се измерват различни характеристики за всяка точка от траекторията. Данните за тези характеристики са в “суров” вид. С прилагането на някои операции към тях (т.нар. предварителна обработка), стойностите на данните се трансформират, за да може да се улесни извличането на признаците от тях.

Координатите  $x$  и  $y$  от координатната система на таблета (*ink coordinate space*) се наричат *himetric units* [28] и приемат високи стойности в  $[0,7999] \times [0,5999]$ . Първата извършвана операция е (1) *трансформиране* на тези координати към  $x$  и  $y$  координатите (нар. пикселови координати) от координатната система на приложението със стойности в  $[0,1279] \times [0,799]$ . Това се извършва автоматично от метод, достъпен от пакета за разработка на приложения Microsoft Tablet PC. При събирането на подписи с графичен таблет има възможност подписите от даден участник да не са подравнени по отношение на оста  $x$ , а да са завъртени на определен ъгъл. Това от своя страна налага нуждата от извършването на (2) *ротация*. Другата извършвана операция е (3) *транслация* на подписите от всички участници в избрана точка от координатната система на приложението.

### Ротация на подпис на определен ъгъл

За извършването на ротацията на подпис се прилага подхода на собствените вектори. Първоначално се намират главните оси на подписа, съответстващи на собствените вектори  $v_1$  и  $v_2$  на ковариационната матрица  $\Sigma_{xy}$ . Пресмятат се и съответните собствени стойности  $\lambda_1$  и  $\lambda_2$ . Максималната собствена стойност  $\lambda_{\max} = \max(\lambda_1, \lambda_2)$  показва оста на най-голямата промяна, а ъгъла  $\theta$ , който подписа сключва с оста  $x$  може да се намери от компонентите на съответния ѝ собствен вектор  $v_{\max}$ .

$$\theta = \arctg\left(\frac{v_{\max 2}}{v_{\max 1}}\right) \quad (2.1)$$

За даден подпис  $S$  с  $n$  на брой точки, представен във вида:

$$S = \{S_i(x_i, y_i), i = 1, \dots, n\} \quad (2.2)$$

завъртането на ъгъл  $\theta$  (докато главната ос съвпадне с тази на оста  $x$ ) се задава със следното матрично уравнение:

$$\begin{pmatrix} x'_i \\ y'_i \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix}, i = 1, \dots, n \quad (2.3)$$

Траекторията от точки  $S' = \{S'_i = (x'_i, y'_i), i = 1, \dots, n\}$  представя подписа след ротация.

### Транслация на подпис в избрана начална точка

За даден подпис  $S$  с  $n$  на брой точки, представен във вида (2.2) и избрана начална точка с координати  $(x_o, y_o)$  първоначално се намират съответните компоненти на транслация  $T_x$  и  $T_y$ :

$$\begin{pmatrix} T_x \\ T_y \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} - \begin{pmatrix} x_o \\ y_o \end{pmatrix} \quad (2.4)$$

след което се извършва самата транслация, като за всяка двойка координати  $(x_i, y_i)$  от стойностите на  $x_i$  и  $y_i$  се извадят съответно стойностите  $T_x$  и  $T_y$ :

$$\begin{pmatrix} x_i^{Tr} \\ y_i^{Tr} \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \end{pmatrix} - \begin{pmatrix} T_x \\ T_y \end{pmatrix}, i = 1, \dots, n \quad (2.5)$$

В резултат на трансформацията, траекторията от точки  $S^{Tr} = \{S_i^{Tr}(x_i^{Tr}, y_i^{Tr}), i = 1, \dots, n\}$  представлява подписа след транслация.

## 2.2 Извличане на признаци на он-лайн подпис от графичен таблет

С помощта на пакета за разработка на приложения *Tablet PC SDK 1.7* се получава динамична информация за събираните с графичния таблет подписи. Тази информация е структурирана в пакети (по един за всяка точка от подписа) и съдържа множество от данни, което се поддържа от съответния графичен таблет. Някои от тези данни са следните:  $X$  и  $Y$  координати, натиск, наклон на писалката, времеви отпечатък (*timestamp*) и азимут. Координатите определят местоположението на върха на писеца

върху координатната система на таблета, натискът приема стойности в интервала  $[0,1023]$  и представлява приложения от писалката натиск в съответната точка. Наклонът на писалката е ъгъла, който тя сключва с повърхността на таблета и приема стойности в интервала  $[0,90]$ , а азимутът е ротацията по часовниковата стрелка на писалката в равнината  $XY$  и приема стойности в интервала  $[0,360]$ .

Таблица 2.1 Глобални признаци

| №   | Признак                                                     | №   | Признак                                                       |
|-----|-------------------------------------------------------------|-----|---------------------------------------------------------------|
| A1  | Дължина на подписа $L$                                      | A14 | Разстояние от най-лявата точка до центъра                     |
| A2  | Височина $H$                                                | A15 | Разстояние от центъра до най-дясната точка                    |
| A3  | Сплеснатост $H/L$                                           | A16 | Ъгъл на отсечката, определена от центъра и най-лявата точка   |
| A4  | Брой точки $N$                                              | A17 | Ъгъл на отсечката, определена от центъра и най-дясната точка  |
| A5  | Общо време                                                  | A18 | Разстояние от най-лявата точка до началната                   |
| A6  | Брой сегменти                                               | A19 | Разстояние от най-дясната точка до крайната                   |
| A7  | Плътност $A4/A1 \cdot A2$                                   | A20 | Ъгъл на отсечката, определена от най-лявата точка и началната |
| A8  | Разстояние от началната точка до центъра                    | A21 | Ъгъл на отсечката, определена от крайната и най-дясната точка |
| A9  | Разстояние от крайната точка до центъра                     | A22 | Брой шрихи                                                    |
| A10 | Разстояние от началната до крайната точка                   | A23 | Наклон на подписа                                             |
| A11 | Ъгъл на отсечката, определена от центъра и началната точка  | A24 | Средна стойност на натиск                                     |
| A12 | Ъгъл на отсечката, определена от центъра и крайната точка   | A25 | Средна стойност на наклон на писалката                        |
| A13 | Ъгъл на отсечката, определена от началната и крайната точка | A26 | Средна стойност на азимут                                     |

Стойностите на признаците за даден подпис се изчисляват въз основа на неговата динамичната информация. В Таблица 2.1 са представени използваните глобални признаци, като във формулите им са използвани следните означения.

Нека с  $S_i$  сме означили  $i$ -тата точка от даден подпис, с  $n$  – броя на точките в подписа,  $(x_i, y_i)$  са координатите на точката  $S_i$ . Натискът в точката  $S_i$  означаваме с  $p_i$ , с  $s_i$  – наклонът в  $i$ -тата точка и с  $t_i$  – изминалото време между изчертаването на  $i$ -тата и  $(i+1)$ -вата точка. С  $(\bar{x}, \bar{y})$  сме означили координати на центъра, с  $(x_{lev}, y_{lev})$  и с  $(x_{desen}, y_{desen})$  сме означили координати на най-лявата и най-дясната точка на подписа.

Центърът  $(\bar{x}, \bar{y})$  на подпис, съставен от  $n$  на брой точки се задава с формулата:

$$(\bar{x}, \bar{y}) = \left( \frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n y_i \right). \quad (2.6)$$

Разстоянието  $d$  между две точки от подписа  $(x_1, y_1)$  и  $(x_2, y_2)$  се задава с:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \quad (2.7)$$

Ъгълът  $\varphi$  на отсечката от точка  $(x_1, y_1)$  до точка  $(x_2, y_2)$  от подписа се задава с:

$$\varphi = \arctg\left(\frac{(y_2 - y_1)}{(x_2 - x_1)}\right). \quad (2.8)$$

Една точка  $S_i$  е локален екстремум, ако е изпълнено едно от условията (2.9):

$$\begin{aligned} y_{i-\varepsilon} < y_i > y_{i+\varepsilon} \\ y_{i-\varepsilon} > y_i < y_{i+\varepsilon} \\ x_{i-\varepsilon} < x_i > x_{i+\varepsilon} \\ x_{i-\varepsilon} > x_i < x_{i+\varepsilon} \end{aligned} \quad (2.9)$$

Големината на  $\varepsilon$  се определя експериментално, като може да бъде равна на 2,3..., но рядко се използва  $\varepsilon = 5$ . Всяка двойка последователни екстремуми разделя подписа на сегменти.

### 2.3 Минимизиране на броя на признаците

В този раздел е представен метод за избор на регресионни променливи, базиран на критерият  $C_p$  на *Mallows* [10]. Подборът на подмножество от независими променливи (фактори), които да се включат в регресионния модел, така че той да дава най-надеждна прогноза, е важна задача в регресионния анализ. Този проблем се нарича избор на най-добро подмножество или избор на най-добро регресионно уравнение [4].

#### 2.3.1 Нормиране на признаците

Тъй като всеки от признаците получава стойности в различни интервали се налага те да се нормират. Чрез извършването на нормировка стойностите на данните попадат в обща скала на отчитане и се премахват единиците за измерване. Сред известните техники за нормиране са стандартизацията (нормировка *z-score*) и нормировката *min-max* [14]. В първия случай, стойностите на признаците се трансформират в такива със средна стойност, равна на нула и стандартно отклонение, равно на единица. Новополучената стойност  $f_i^{zscore}$  за признак  $f_i$  се намира по формулата:

$$f_i^{zscore} = \frac{f_i - \mu_{f_i}}{\sigma_{f_i}}, i = 1, \dots, k \quad (2.10)$$

където с  $\mu_{f_i}$  и  $\sigma_{f_i}$  са означени съответно средната стойност и стандартното отклонение за  $i$ -тия признак, пресметнати за всички подписи от съответния потребител, изчислени по формулите:

$$\begin{aligned} \mu_{f_i} &= \frac{1}{n} \sum_{j=1}^n f_{ij} \\ \sigma_{f_i} &= \sqrt{\frac{1}{n-1} \sum_{j=1}^n (f_{ij} - \mu_{f_i})^2} \end{aligned} \quad (2.11)$$

Нормировката *min-max* в интервала  $[c, d]$  се извършва за стойностите на всеки признак по формулата:

$$f_i^{\min \max} = \frac{f_i - \min_{f_i}}{\max_{f_i} - \min_{f_i}} (d - c) + c, i = 1, \dots, k \quad (2.12)$$

където с  $\min_{f_i}$  и  $\max_{f_i}$  са означени съответно минимума и максимума за  $i$ -тия признак, пресметнати по всички подписи от съответния потребител.

$$\min_{f_i} = \min_j (f_{ij}) \quad (2.13)$$

$$\max_{f_i} = \max_j (f_{ij})$$

В задачите за разпознаване на образи нормировката най-често се извършва в интервала  $[0,1]$ .

### 2.3.2 Премахване на един от група признаци с висока корелация

Корелацията е мярка за измерване както на посоката, така и на силата на връзката между две променливи [16]. Стойността на корелация се определя от стойността на корелационния коефициент. Най-често използван е корелационния коефициент  $r$  на Pearson, който за два  $p$ -мерни вектора  $\mathbf{X}_1$  и  $\mathbf{X}_2$  се определя по формулата:

$$r_{X_1, X_2} = \frac{\sum_{i=1}^p (X_{1i} - \mu_{X_1})(X_{2i} - \mu_{X_2})}{\sqrt{\sum_{i=1}^p (X_{1i} - \mu_{X_1})^2 \sum_{i=1}^p (X_{2i} - \mu_{X_2})^2}} \quad (2.14)$$

където с  $\mu_{X_j}$  е означено средното за вектора  $\mathbf{X}_j$ ,  $j=1,2$

$$\mu_{X_j} = \sum_{i=1}^p X_{ji}, j=1,2 \quad (2.15)$$

Големината на корелационния коефициент е показател за силата на линейната връзка между векторите  $\mathbf{X}_1$  и  $\mathbf{X}_2$ . Колкото той е по-близко до 1 или -1, толкова е по-силна връзката между двата вектора. Стойността на коефициента на корелация би могла в голяма степен да се повлияе от един или повече признаци, силно отдалечени от групата данни (*outliers*) [16].

Корелационната матрица на признаците с размер  $k \times k$  може да се представи по следния начин за даден потребител  $m = 1, \dots, M$

$$C_{k \times k}^m = \{C_{ij}^m\}, i, j = 1, \dots, k \quad (2.16)$$

Всеки не диагонален елемент  $C_{ij}^m$ ,  $i \neq j$  на корелационната матрица представлява стойността на коефициента на корелация между признака  $i$  и признака  $j$  за съответния потребител  $m$ .

Методът на корелационните плеяди [30] се състои в намирането на групи от признаци с висока стойност на корелация между признаците в групата и ниска стойност на корелация между признаците от отделните групи. Тези групи признаци се наричат корелационни плеяди. С прилагането на този метод се намалява броя на признаците като от всяка група се оставя единствен признак. Няма теоретически обосновано становище кой признак от плеядата да бъде запазен. Най-простият и практически оправдан подход е да се избере този признак, получаването на който е най-бързо и лесно. Значително по-убедително би било, ако за всяка комбинация от избрани признаци се оцени точността на класификацията. Това обаче изисква значителни изчислителни усилия, а резултатът от практическа гладна точка едва ли ще бъде съществен точно поради силната корелационна връзка в плеядата.

### 2.3.3 Избор на най-информативни признаци

Регресионният анализ е често използван метод за установяване на функционална зависимост между независимите променливи и зависимата променлива. Намира широко приложение в различни приложни задачи [4]. Изборът на регресионни променливи в регресионния анализ се състои в определянето на подмножество от променливи, които да се включат в модела, така че той да дава надеждна прогноза за принадлежността на дадена точка към коректния клас.

Формулировката на задачата за избор на регресионни променливи е следната [14]. Нека са дадени  $n \geq k+1$  наблюдения над  $k$ -мерния вектор от независими променливи  $\mathbf{x}_j^t = (x_{1j}, \dots, x_{kj})$  и зависимата променлива  $y_j$ ,  $j = 1, \dots, n$ , стойността на която се определя от:

$$y_j = \sum_{i=1}^k \beta_i x_{ij} + e_j \quad (2.17)$$

Моделът на регресията (10) има следния матричен вид:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (2.18)$$

където  $\mathbf{Y}$  е  $n$ -мерен вектор на зависимите променливи,  $\mathbf{X}$  е матрица на независимите променливи с размерност  $n \times k$ ,  $\boldsymbol{\beta}$  е  $k$ -мерен вектор на неизвестните регресионни коефициенти и  $\mathbf{e}$  е  $n$ -мерен вектор на остатъците, за които се предполага, че са независими и нормално разпределени  $N(0, \sigma^2)$  с математическо очакване, равно на нула и дисперсия, равна на  $\sigma^2$ . Регресионните параметри (коефициенти)  $\beta_i$ ,  $i = 1, \dots, k$  са неизвестни константи, които следва да се оценят по метода на най-малките квадрати от изходните данни. Техните оценки се означават с  $b_i$ . Тогава, векторът  $\mathbf{b}$  от оценките на вектора на коефициентите по метода на най-малките квадрати се задава чрез:

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (2.19)$$

Най-популярният подход за избор на регресионен модел е този за “избор на най-добро подмножество” (*best subset*). При прилагането му се разглеждат всичките  $2^k - 1$  на брой подмножества от регресионни променливи. Това на практика е неприложимо, особено ако броят на променливите е голям. В литературата съществуват методи, при които търсенето на подмножествата от променливи не включва пресмятания върху цялото признаково подмножество, а само върху част от него [15]. Към тях се отнася и разглеждания по-долу метод на Hocking и Leslie. Повечето от критериите за избор на регресионни променливи са монотонни функции на остатъчната сума от квадрати (*RSS*, *residual sum of squares*) и поради тази причина задачата за избор на регресионни променливи се редуцира до намирането на тези подмножества от променливи, за които остатъчната сума от квадрати е малка. Остатъчната сума от квадрати  $RSS_p$  за  $n$  наблюдения над  $p$  на брой променливи се дефинира като:

$$RSS_p = \sum_{j=1}^n (y_j - \sum_{i=1}^p \beta_i x_{ij})^2 \quad (2.20)$$

*Mallows* [17] предлага критерий за сравняване на няколко регресионни модела с цел откриването на най-добрите от тях. За целта се пресмята стойността на  $C_p$  за всеки

един от моделите. За модел с  $p$  на брой променливи и  $n$  наблюдения,  $C_p$  се дефинира като:

$$C_p = \frac{RSS_p}{\hat{\sigma}^2} - (n - 2p) \quad (2.21)$$

Оценката  $\hat{\sigma}^2$  на  $\sigma^2$  се задава по формулата:

$$\hat{\sigma}^2 = \frac{1}{n - k} \sum_{j=1}^n (y_j - \sum_{i=1}^k \beta_i x_{ij})^2 = \frac{RSS_k}{n - k}. \quad (2.22)$$

където с  $\sigma^2$  е означена дисперсията на  $n$ -мерния вектор на остатъците  $\mathbf{e}$ .

Така, ако даден модел с  $p$  на брой променливи е добър, то от (2.21) и (2.22) се намират следните стойности за математическото очакване на стойностите на  $RSS_p$  и  $C_p$

$$\begin{aligned} E(RSS_p) &= (n - p)\hat{\sigma}^2 \\ E(C_p) &\approx p \end{aligned} \quad (2.24)$$

За добри подмножества от променливи се считат тези с малки стойности на  $C_p$  или със стойности на  $C_p$ , близки до  $p$  [17].

Hocking и Leslie [15] създават метод, който позволява откриването на най-добри подмножества от регресионни променливи след разглеждането на малка част от всичките  $\binom{k}{p}$  възможни подмножества с размер  $p$ . LaMotte и Hocking [9] модифицират метода така, че сравнително големи по обем задачи могат да се решават с минимум изчисления. С т.нар.  $r$ -подмножество се означава подмножеството от променливи, които се изключват от регресионния модел, а с  $p$ -подмножество – подмножеството от променливи, които остават в модела. По същество методът се базира на т.нар. редукции на  $m$ -подмножества, т.е. на редукции в остатъчните суми от квадрати поради премахването на подмножества с размер  $1 \leq m \leq 4$  от модела с  $k$  на брой променливи. Редукциите на  $m$ -подмножества служат за определянето на най-добро  $r$ -подмножество, което да се премахне (за  $r > m$ ). Редукцията на регресионната сума от квадрати  $Red_r$ , получена от премахването на подмножество от  $r$  променливи се задава чрез:

$$Red_r = RSS_p - RSS_k \quad (2.25)$$

Множеството от  $r$  променливи, за което стойността на тази редукция е най-малка определя подмножеството от  $p$  ( $p = k - r$ ) регресионни променливи, за което остатъчната сума от квадрати е минимална. В [15] е предложена следната формула за намиране на  $C_p$  чрез тази редукция:

$$C_p = \frac{Red_r}{\hat{\sigma}^2} - (2p - k) \quad (2.26)$$

По-долу са изброени стъпките на обобщения алгоритъм [9].

1. Намират се регресионните коефициенти за модела с всичките  $k$  на брой променливи;
2. За всяка от променливите  $x_i$  се пресмята редукцията  $\theta_i$  от премахването ѝ:

$$\theta_i = \hat{\sigma}^2 t_i^2 \quad (2.27)$$

където  $t_i$  е  $t$ -статистиката на  $i$ -тия регресионен коефициент.

Процедурата продължава с пренареждане на променливите по нарастване на съответните им редукции:

$$\theta_1 \leq \theta_2 \leq \dots \leq \theta_k$$

Методът се основава на следното фундаментално свойство на квадратичните форми: “Ако редукцията на регресионната сума от квадрати, получена от премахването на подмножество от променливи с най-голям индекс  $j$  не надвишава стойността на редукцията  $\theta_{j+1}$  от премахването на  $(j+1)$ -вата променлива, то не съществува подмножество, включващо променливи с индекси, по-големи от  $j$ , за което редукцията да е по-малка”.

3. Пресмятат се редукциите на  $m$ -подмножества за всяко от всичките  $\binom{k}{m}$  на брой подмножества с размер  $m$ . Всяко от тези подмножества с размер  $m$  се представя във вида  $(i_1, i_2, \dots, i_m)$ , където  $i_1, i_2, \dots, i_m$  са променливите, от които е съставено, подредени в нарастващ ред,  $1 < i_j < k$  и  $i_1 < i_2 < \dots < i_m$ . Така пресметнатите редукции на  $m$ -подмножества се подреждат в нарастващ ред. Тези  $m$ -подмножества, чиито първи индекс е по-голям или равен на  $(r - m + 1)$  служат за дефиниране на етапи за изследване на  $r$ -подмножества. Подмножествата с размер  $r$ , определени на даден етап съдържат  $m$  на брой индекси от съответното  $m$ -подмножество и  $(r - m)$  индекси, по-малки от първия индекс в съответното  $m$ -подмножество. Броят на дефинираните по този начин етапи е  $\binom{k-r+m}{m}$ .

Етапите се номерират според големината на стойността на редукцията на  $m$ -подмножествата, които ги дефинират. Така, първият етап се състои от  $r$ -подмножествата, определени от  $m$ -подмножеството с най-малка стойност на редукция.

4. Изобщо, на всеки етап  $q$  се прави проверка дали най-малката от редукциите на  $r$ -подмножествата, дефинирани от всички етапи до момента  $(1, \dots, q)$  е по-малка от редукцията на  $m$ -подмножеството, дефиниращо следващия  $(q+1)$ -ви етап, и ако това е изпълнено, се счита, че е открито търсеното  $r$ -подмножество, което да напусне изходния модел и съответстващото му най-добро  $p$ -подмножество, а в случай, че не е изпълнено, се преминава към следващия  $(q+1)$ -ви етап.



## 2.4 Резултати в Глава 2

1. Формулирани са изискванията към използваните за верификация признаци.
2. Описани са трансформациите на получените от графичния таблет изображения на подписи, необходими за привеждането им в стандартен вид и извличането на признаци. Представени са 26 глобални признака, необходими за математическото описание на подписа. Признаците с номера A8-A21 не се срещат в литературата до момента.
3. Описан е подход, основаващ се на последователното прилагане на метода на корелационните плеади и  $C_p$  статистиката на *Mallows*, за избор на най-информативно от гледна точка на класификацията индивидуално за всеки потребител подмножество от признаци.

## Глава 3. Класификация

Тази глава обхваща следните раздели.

### 3.1 Невронните мрежи като класификатори

#### 3.1.1 Видове невронни мрежи

#### 3.1.2 Обучение на невронна мрежа

#### 3.1.3 Предимства и недостатъци на невронните мрежи

#### 3.1.4 Невронна мрежа за класификация

#### 3.1.5 Приложение на невронните мрежи в задачата за разпознаването по подпис

Невронните мрежи са подходящо средство за решаване на сложни задачи като разпознаването на подпис поради следните причини [13]. На първо място, в случай че са добре обучени, те обикновено дават верни отговори за входни образци, неизползвани при обучението. Също така многослойните невронни мрежи могат да моделират всяка функция на множество от променливи, а подписите могат лесно да се представят като такава функция. Поради възможността им за обобщаване, невронните мрежи могат да се справят с разнообразието и изменчивостта, характерни за образците на подписите. Тъй като в повечето случаи обучението на невронна мрежа отнема дълго време, в приложенията за разпознаване по подпис то обикновено се извършва еднократно и в оф-лайн режим, като по този начин на потребителите се спестява неудобството да чакат до приключването му. През последните години в литературата са предложени няколко системи за разпознаване на подпис, създадени чрез прилагането на невронни мрежи [13]. Идеалното обучение на една невронна мрежа за разпознаване по подпис е практически невъзможно, поради нестабилния характер на данните (подписът на даден потребител зависи от възрастта му, от неговото физическото и психическо състояние, от употребата на медикаменти и опиати, от вида на хартията, върху която се подписва, от вида на писалката и още много други).

Съществуват два режима за реализация на система за разпознаване на подписи с невронни мрежи, а именно верификация и идентификация. За режим верификация са необходими както собствени, така и чужди подписи от всеки потребител. За всеки участник се създава и съхранява единствена невронна мрежа, която има само един изходен неврон, чийто краен отговор е ниво на достоверност, приемащо стойности между нула (чужд подпис) и единица (собствен подпис), която показва степента на сходство на тестовия подпис с подписите, използвани при обучението за съответния

участник. Тази стойност се сравнява с предварително избрана стойност на прага за участника и ако я надвишава, тестовият подпис се счита за собствен, а в противен случай за чужд. Този подход е широко приложим на практика и позволява бързото добавяне и премахване на подписи на нови и стари потребители. Така, при добавяне на нов потребител е необходимо единствено да се обучи малка по размер невронна мрежа.

За реализацията на режим идентификация са необходими единствено собствени подписи от всеки потребител. В този случай се създава и съхранява една обща невронна мрежа, която има толкова изходни неврони, колкото е броя на участниците. Изходът от идентификацията са толкова на брой стойности между нула и единица, колкото е и броят на всички потребители и като резултат се избира участника, съответстващ на максималната изходна стойност на мрежата. Когато се използва обща невронна мрежа, е необходимо тя да се обучи наново при добавяне на участник, което несъмнено изисква повече време.

### 3.1.6 Проектиране на невронни мрежи

Задачата за проектиране на невронна мрежа включва следните подзадачи : *събиране на данни, създаване на мрежата, конфигуриране на мрежата, инициализиране на теглата и отклоненията ѝ, обучение, оценка на мрежата и същинското ѝ приложение* [11].

- 1) *Събиране на данни* - Качеството на събраните данни и множеството от наличните входни данни (признаци), които постъпват на входа ѝ са от голямо значение.
- 2) *Създаване на мрежата* - Броят на елементите от входния слой зависи от размерността на входните данни, а броят на елементите от изходния слой - от броя на разпознаваните класове. Добра практика е [25] да се изберат некорелирани входни променливи и да се извърши тяхната нормиране.
- 3) *Конфигуриране на мрежата* - При проектирането на невронната мрежа следва да се вземе решение относно нейната топология (брой на слоевете и броя на невроните във всеки слой, техния тип и свързаност) и нейните параметри. Те са силно зависими от решавания проблем и наличните данни и затова стойностите им се намират експериментално.
- 4) *Инициализиране на теглата и отклоненията на мрежата* - Тъй като успехът на проектираната НМ зависи от произволните стойности на началните тегла, е необходимо да се обучат няколко НМ с различни инициализации на теглата и да се избере тази, която демонстрира най-добро качество.
- 5) *Обучение* - След като топологията на мрежата е дефинирана, е необходимо да се избере вида на обучаващия алгоритъм и да се премине към нейното обучение, така че да се оптимизира качеството ѝ.
- 6) *Оценка на мрежата и приложение* – За оценяване на мрежата може да служи оценката на грешката *PCTError*, отразяваща процентното съотношение на грешно класифицираните образци към общия брой образци.

### 3.2 Метод на $k$ -най-близки съседи

Методът на  $k$ -най-близки съседи е непараметричен метод за класификация. Отношение към точността на класификацията по метода на  $k$ -най-близки съседи имат: (1) стойността на  $k$ , (2) избора на мярка за разстояние или сходство, както и (3) метода за комбинирането на етикетите на класовете. В случай на класификация в два класа (дихотомия) се препоръчва нечетна стойност на  $k$ . Максимална стойност за  $k$  е  $k_{max} = \sqrt{n}$  [13]. За определяне на стойността на  $k$  е удачно да се оцени точността на метода чрез LOOCV за множество от стойности на  $k$  и да се избере тази от тях, която води до най-висока точност [6].

### 3.3 Оценяване и сравняване на качеството на класификатори

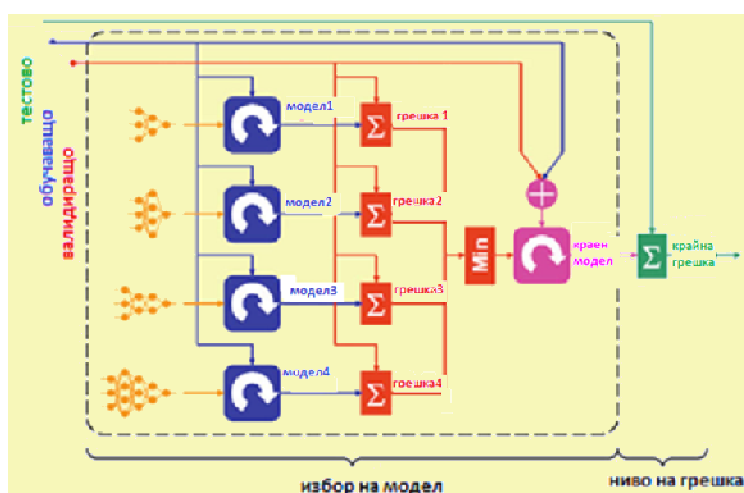
След проектирането на даден класификатор (в частност невронни мрежи) е необходимо да се оцени неговото качество. Оценката на точността на класификатори, обучени с учител, е важна не само за предсказване на точността при бъдещото им използване, но също така и при избор на класификатор и комбинирането на класификатори. Традиционната практика е множеството от проектираните модели да се оценят и измежду тях да се избере този с най-малка грешка.

#### 3.3.1 Оценка на класификатори

В случая на разпознаване в два класа, съответно *Positive* (положителни образци) и *Negative* (положителни образци) се намира стойност на праг, който да ги разграничава най-добре. Така, всеки един пример попада в даден клас в зависимост от това дали резултата от разпознаването надхвърля стойността на прага (клас *Positive*) или не (клас *Negative*) [29]. Това разпределяне би могло да бъде вярно (*True*) или грешно (*False*). Така, за даден образец от клас *Positive*, ако изходът от класификацията е положителен, то образецът се явява истински положителен (*True Positive, TP*), а ако е отрицателен, то образецът е лъжливо отрицателен (*False Negative, FN*). Аналогично, даден образец от клас *Negative* би могъл да бъде истински отрицателен (*True Negative, TN*) или лъжливо положителен (*False Positive, FP*). Въз основа на *TP*, *FN*, *TN* и *FP* са дефинирани следните мерки за оценка на качеството на класификатори: *FAR*, *TAR*, *FRR*, *TRR*, *AC*, *Precision*, *PCTError*, *Sensitivity*, *Specifity*. За нагледно изобразяване на качеството на биометричните системи обикновено се изчертават т.нар. *Receiver Operating Characteristic* (ROC) криви, които представляват визуални средства за оценяване на точността на класификатор и за сравняване на различни класификационни модели.

#### 3.3.2 Метод Holdout

Методът *Holdout* се прилага при наличието на достатъчно примери за обучение. Множеството от образци се разделя в две извадки: за обучение и за тестване. Образците от обучаващата извадка се използват за определяне на параметрите на модела, а тези от тестовата извадка се използват за получаването на оценка на качеството на класификатора върху непознати и неизползвани за обучението му данни и за сравняването на различни мрежови архитектури.



Фигура 3.2 Метод Holdout

Прилагането на метода *Holdout* в случая на невронни мрежи (Фигура 3.2) се състои в следното. Образците се разделят в [11]: обучаваща, валидираща и тестова

извадки. Данните от обучаващата извадка се представят по време на обучението и се използват за настройване на теглата на невронните мрежи. Тези от валидиращата извадка се използват за измерване на качеството на всяка от проектираните невронни мрежи върху непознати данни и при взимането на решение за прекратяване на нейното обучение, когато качеството престане да се подобрява. Данните от тестовата извадка се използват за симулиране на невронната мрежа и измерване на нейното качество след приключване на обучението ѝ. Така кандидат невронните мрежи се нареждат във възходящ ред по качеството им, изчислено върху валидиращата извадка, а крайната грешка се смята върху данните от тестовата извадка.

### 3.3.3 Метод на $N$ -кратна крос валидация ( $N$ -fold cross validation)

Методът на  $N$ -кратна крос валидация е приложим в случаите, когато на разположение са малък брой примери за обучение. Същността на метода се състои в следното. Данните се разделят произволно на  $N$  части (*folds*) с приблизително равен брой образци във всяка част. Последователно се съставят всичките  $N$  на брой възможни извадки, всяка от които съдържа обучаваща извадка, състояща се от  $(N-1)$  части и тестова извадка, съдържаща останалата една част. По този начин всяка част се използва за тестване точно един път. С усредняване на стойностите на грешките, получени за всяко разделяне на данните, се намира оценка на стойността на крайната грешка на класификатора.



Фигура 3.4 Метод на  $N$ -кратна крос-валидация

Характерно за този метод е това, че за обучението на класификатор се използват всичките налични примери за обучение и по този начин не се губят никакви данни. Методът на  $N$ -кратна крос валидация е изобразен схематично на Фигура 3.4.

Ако стойността на  $N$  е равна на размера на обучаващата извадка, то методът се нарича *Leave One Out Cross Validation (LOOCV)*. В този случай, тестовата извадка се състои от единствен образец, а за обучението се използват възможно най-много данни.

С прилагане на метода на  $N$ -кратна крос валидация може да се провери способността за обобщаване на невронната мрежа [11, 12] в случаите, когато на разположение са малък брой примери за обучение.

Невронната мрежа се обучава за всяко от тези  $N$  на брой множества с  $K$  на брой инициализации на началните тегла на връзките. По този начин за всяко разделяне  $i = 1, \dots, N$  се намират  $K$  на брой оценки на грешките  $e_{i,j}$   $j = 1, \dots, K$ , а стойността на крайната грешка за всяко разделяне  $E_i$  се изчислява като средната стойност от тези  $K$  на брой оценки:

$$E_i = \frac{1}{K} \sum_{j=1}^K e_{i,j}, \quad i = 1, \dots, N; j = 1, \dots, K \quad (3.13)$$

Оценките  $E_i$  на съответните класификатори са неизместени, тъй като данните от тестовата извадка не се съдържат в обучаващата извадка. Оценката на конкретния модел  $\hat{E}$  се намира като средната стойност от грешките, получени за всяко разбиране на данните:

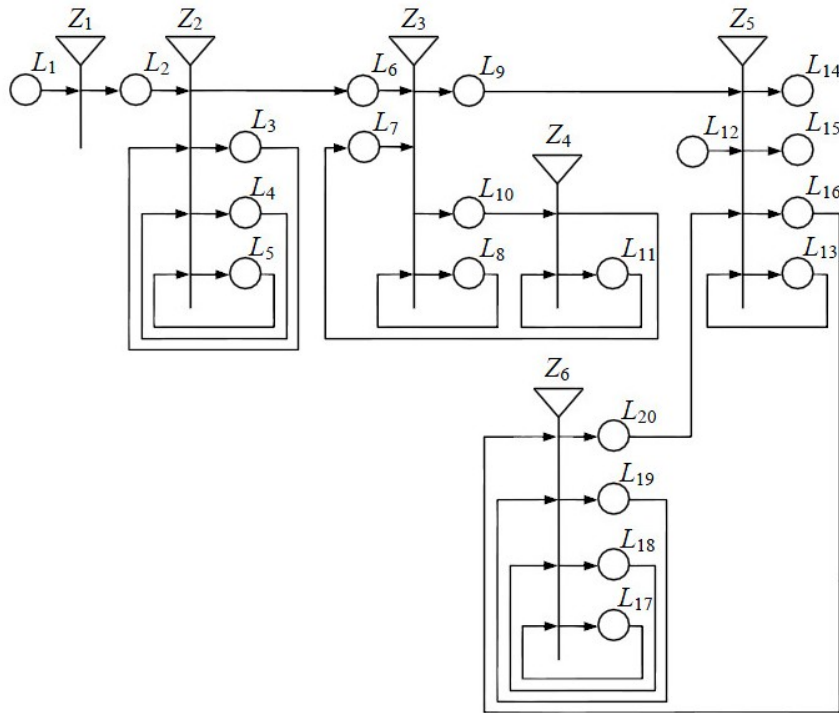
$$\hat{E} = \frac{1}{N} \sum_{i=1}^N E_i \quad (3.14)$$

Оценката  $\hat{E}$  разчита на допускането, че точността на класификацията би се променила плавно с промяна в размера на обучаващата извадка. Следователно, ако точността на модела е задоволителна, изследователят би могъл или да обучи единствен класификатор върху цялото множество от данни, или да комбинира класификаторите, построени по време на обучението [27].

Методът на  $N$ -кратна крос валидация е удачен избор при търсенето на подходяща стойност на  $k$  в класификатора  $kNN$ , а също и при търсенето на оптимална топология на НМ [6]. След намирането на оптимални стойности на параметрите на класификатор, той се обучава върху цялото множество от данни и бива използван на практика.

### 3.4 Обобщен мрежови модел за он-лайн верификация на подписи

С апарата на обобщените мрежи (ОМ) е направен опит за моделиране на процеса на верификация по подпис с цел ясно и нагледно представяне на предложения в дисертацията метод, който включва следните стъпки за верификация на заявена самоличност по подпис: полагане на подпис, предварителна обработка на подпис, извличане на признаци, избор на признаци, обучение и верификация.



Фигура 3.6 Обобщено мрежови модел на процеса

Представената редуцирана обобщена мрежа е без времеви компоненти, без зададени приоритети на преходите, позициите и ядрата и представлява модификация на предложените в [35, 36] модели на ОМ за различни задачи от разпознаването на образи. Обобщено мрежовият модел е представен на Фигура 3.6. Той е изграден от 6 прехода (всеки един от които представя стъпка от процеса), 20 позиции и 3 вида ядра  $\alpha$ ,  $\beta$  и  $\gamma$ .

### 3.5 Резултати в Глава 3

1. Представена е кратка класификация на различни видове невронни мрежи според архитектурата и алгоритмите за обучение. Коментирани са предимствата и недостатъците им.
2. Анализирани са въпросите за изграждане и обучение на невронни мрежи.
3. Описани са видовете грешки от класификация, методите за оценка на точността и построяването на комбиниран класификатор.
4. Изграден е модел на процеса на он-лайн верификация с апарата на обобщените мрежи.

## Глава 4. Експерименти и резултати

В настоящата глава са представени получените резултати от провеждането на експериментите за създаване и тестване на комбинираната система за разпознаване по подпис. Тази система може да служи за верификация въз основа на признаците, описани в Раздел 2.2. За решаване на поставената задача са проектирани класификатори невронни мрежи от тип многослоен перцептрон и класификатори по  $k$ -най-близки съседи, които са тествани върху образците на подписи от две бази данни. Комбинираният подход за разпознаване по подпис се прилага като поредица от следните стъпки: (1) предварителен избор на признаци и (2) верификация.

### 4.1 Избор на признаци за разпознаване

За всеки от потребителите се извличат стойностите на признаците, изчислени по характеристиките на всичките му осем или десет собствени и десет подправени подписа. В настоящата дисертация е експериментирано с умели и неумели фалшификати. Неумелите фалшификати са избрани произволно от собствените подписи на останалите участници. Признаците са глобални и са описани в раздел 2.2. Тъй като всеки от тях получава стойности в различни интервали при различна скала на измерване, се налага необходимостта от нормиране на признаците в интервала  $[0,1]$ . Извършва се нормировка *min-max* (раздел 2.3.1), която запазва взаимовръзките между оригиналните стойности на данните. След извличането на признаци, се пристъпва към намирането на (1) **общо** за всички потребители подмножество от признаци с прилагане на метода на корелационните плеяди (точка 2.3.2); (2) **индивидуално** подмножество от признаци на базата на собствени подписи и **неумели** фалшификати (наричано по-долу Вариант “1”); (3) **индивидуално** подмножество от признаци на базата на собствени подписи и **умели** фалшификати (наричано по-долу Вариант “2”). Така намерените подмножества от признаци служат за построяването на класификационните модели.

След намирането на общо за всички потребители подмножество от признаци, се прилага метода за избор на регресионни променливи, описан в Раздел 2.3 за намирането на индивидуалните подмножества от признаци (2) и (3). Така се откриват най-добрите подмножества от признаци с различен размер за всеки участник въз основа на неговите собствени и подправени подписи. Сред тях се избира това подмножество от признаци, за което стойността на  $S_r$  е най-близка до тази на  $p$ , където  $p$  е броят на регресионните коефициенти. Получените по този начин подмножества от признаци са с

различен размер за различните потребители. Нека означим броя на признаците с  $k$ , а този на подписите на даден потребител с  $n$ . За всеки потребител се създава текстов файл с двадесет реда, равен на броя подписи, определени за обучение. Всеки ред съдържа стойности на признаците за съответния подпис, разделени с точка и запетая, следвани от 1 (ако конкретният подпис е собствен, тоест принадлежи на първия клас) и -1 (ако подписът е подправен, тоест принадлежи на втория клас). Необходимо е да се подберат тези признаци (независими променливи), които възможно най-добре апроксимират наблюдаваните значения на зависимата променлива  $y$  ( $y = 1$  за точките от първия клас, собствен подпис и  $y = -1$  за тези от втория клас, фалшифициран подпис). Този файл се обработва и за всеки потребител / файл се откриват най-добрите подмножества с размер  $p$  ( $p = k - r$ ) за стойности на  $r$  ( $3 < r < 13$ ), а сред тях се намира подмножеството със стойност на  $C_p$  най-близка до  $p$ , но по-малка от нея. Това подмножество от признаци се приема за “най-добро”.

## 4.2 Построяване на модели за класификация

Експериментите са извършени с класификатори невронни мрежи и класификатор по  $k$ -най-близки съседи. Построени са класификатори с вариращи параметри (брой признаци, вид фалшифицирани подписи за обучение, стойност на  $H$  при НМ и стойност на  $k$  при  $kNN$ ). Поради малкия обем данни за обучение, точността на различните модели за класификация е оценена с 10-кратна крос валидация (НМ) и LOOCV ( $kNN$ ).

### 4.2.1 Класификатор невронна мрежа

За всеки потребител се създава отделна НМ. Входът  $p$  на всяка от тестваните невронни мрежи представлява матрица [*FeatureSize Ntrn*] с брой редове, равен на броя на признаците *FeatureSize* и брой стълбове, равен на броя на образци за обучение *Ntrn*. Изходът зависи от режима на разпознаване. В случай на верификация той е матрица с размерност [*1 Ntrn*], със стойност в  $[0,1]$  във всеки от редовете, равна на 1 за собствен подпис и 0 – за подправен.

В процеса на обучение участват и собствени и подправени подписи. Експериментирано е с използването само на неумели фалшификати (наричано по-долу Случай ”1”) и едновременно на умели и неумели фалшификати (наричано по-долу Случай ”2”). В търсене на оптимална архитектура на НМ, за всеки потребител се експериментира с различни топологии на един и същ вид невронна мрежа според променливия размер на броя на невроните в скрития слой. Топологията на НМ се избира чрез крос валидация.

Така, за всеки потребител се намират параметри на модел на НМ (брой скрити неврони, вид фалшифицирани подписи за обучение, брой признаци на входа и др.). Изходът от верификацията е стойност от интервала  $[0,1]$ . Окончателната невронна мрежа се запазват заедно с архитектурата и стойностите на теглата и може да се използва по-нататък за верификация.

### 4.2.2 Класификатор по $k$ -най-близки съседи

За създаването на класификатор по  $k$ -най-близки съседи са следвани препоръките, изложени в Раздел 3.2. Първоначално се прилага нормиране на признаците, така че те да имат стандартно отклонение, равно на единица и средна стойност, равна на нула. Използвано е Евклидово разстояние за пресмятане на разстоянията между образците. Тъй като  $k_{max} = \sqrt{Ntrn}$ , *Ntrn* – брой примери за обучение и стойността на  $k$  е препоръчително да е нечетно число, то са тествани са две стойности на  $k$ : 1 и 3.

### 4.3 Получени резултати върху собствена база от данни за подписи

Резултатите от избора на най-информативните признаци по метода на Hosking и Leslie (Раздел 2.3) за всеки от осемте потребителя са обобщени в таблица 4.2. От нея е видно, че в случая на вариант “2” броя на признаците за всички потребители без един се редуцира до 13, като само за потребител #5 те са 5 признака, докато при вариант “1” броя на признаците за половината участници се редуцира до брой, по-малък от 13.

Таблица 4.2 Редуцирани признаци за всеки потребител

| Участник | #1  |     | #2   |     | #3  |      | #4  |     | #5  |     | #6   |      | #7   |      | #8   |      |
|----------|-----|-----|------|-----|-----|------|-----|-----|-----|-----|------|------|------|------|------|------|
| Вар.     | “1” | “2” | “1”  | “2” | “1” | “2”  | “1” | “2” | “1” | “2” | “1”  | “2”  | “1”  | “2”  | “1”  | “2”  |
| A2       | +   | +   | +    |     |     | +    |     | +   | +   |     |      | +    | +    | +    | +    |      |
| A5       |     | +   | +    | +   | +   | +    | +   | +   | +   | +   | +    |      | +    | +    |      | +    |
| A6       | +   |     | +    | +   |     | +    |     | +   | +   | +   | +    | +    |      |      | +    | +    |
| A8       |     | +   | +    | +   |     | +    | +   | +   | +   | +   |      | +    | +    |      | +    |      |
| A11      | +   | +   |      |     |     | +    |     |     |     |     | +    |      |      | +    | +    | +    |
| A12      | +   |     |      | +   |     | +    |     | +   | +   |     |      | +    | +    | +    | +    | +    |
| A13      | +   | +   |      | +   | +   | +    |     | +   |     |     | +    | +    | +    | +    |      | +    |
| A15      | +   | +   | +    | +   |     | +    |     | +   |     |     | +    | +    | +    | +    |      |      |
| A16      |     |     | +    | +   | +   | +    |     | +   | +   |     | +    | +    | +    | +    | +    |      |
| A17      |     | +   | +    | +   | +   | +    | +   | +   | +   |     | +    | +    |      | +    |      | +    |
| A20      | +   | +   | +    | +   | +   |      | +   |     | +   |     | +    |      |      | +    |      | +    |
| A21      | +   | +   | +    | +   | +   | +    | +   | +   | +   |     | +    | +    | +    | +    | +    | +    |
| A22      | +   | +   |      |     |     |      |     | +   | +   | +   | +    | +    | +    | +    | +    | +    |
| A23      | +   | +   | +    |     | +   |      | +   |     |     |     | +    |      | +    |      | +    | +    |
| A24      | +   | +   | +    | +   | +   | +    | +   | +   |     |     | +    | +    | +    |      | +    | +    |
| A25      | +   |     | +    | +   |     |      |     |     | +   |     |      | +    | +    | +    |      | +    |
| A26      | +   | +   | +    | +   | +   | +    | +   | +   | +   | +   | +    | +    | +    | +    | +    | +    |
| Ср       | 10, | 9,0 | 9,09 | 9,5 | 8,1 | 9,05 | 7,1 | 9,2 | 10, | 4,8 | 9,08 | 9,33 | 9,12 | 9,24 | 10,3 | 9,77 |
| р        | 13  | 13  | 13   | 13  | 9   | 13   | 8   | 13  | 12  | 5   | 13   | 13   | 13   | 13   | 11   | 13   |

Проектираните невронни мрежи са 30 на брой (Таблица 4.3), а проектираните класификатори по  $k$ -най-близки съседи са 12 на брой за всеки участник.

Таблица 4.3 Параметри на моделите

| # на модел | Брой признаци (входни) | Брой собствени подписи | Брой подправени подписи |            | Брой скрити неврони | Брой съседи $k$ |
|------------|------------------------|------------------------|-------------------------|------------|---------------------|-----------------|
| 1          | Общи признаци          | 10                     | 15 неумели              | Случай “1” | от 1 до 5           | 1 или 3         |
| 2          | Вариант “1”            |                        |                         |            |                     |                 |
| 3          | Вариант “2”            |                        |                         |            |                     |                 |
| 4          | Общи признаци          |                        |                         |            |                     |                 |
| 5          | Вариант “1”            |                        |                         |            |                     |                 |
| 6          | Вариант “2”            |                        |                         |            |                     |                 |

В случай на равни максимални стойности на точността за два или повече модела за даден участник, е избран този от тях, построен с по-малък брой признаци и / или



брой скрити неврони (НМ) и / или брой съседи (kNN). Средната стойност на приближената оценка на точността за класификатор НМ за всички участници е равна на 97.5%, а тази на точността за класификатор kNN е равна на 92.5%.

След намирането на оптимални стойности на параметрите на класификатор по метода на крос валидация, той се обучава върху цялото множество от данни и бива използван на практика.

#### 4.4 Получени резултати върху база данни за подписи SUsig

Резултатите от избора на най-информативни признаци по метода на *Hocking* и *Leslie* (Раздел 2.3) за всеки от потребителите са обобщени в таблица 4.7.

Таблица 4.7 Редуцирани признаци за всеки потребител от SUsig

| Участник | Вар. | A1 | A2 | A4 | A6 | A10 | A12 | A13 | A16 | A17 | A21 | A22 | A23 | A24 | Ср   | p |
|----------|------|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|---|
| #1       | "1"  | +  |    | +  | +  | +   |     | +   | +   |     |     | +   | +   | +   | 5,35 | 9 |
|          | "2"  | +  |    |    | +  |     | +   | +   | +   |     | +   | +   | +   | +   | 5,25 | 9 |
| #2       | "1"  | +  |    |    | +  | +   | +   | +   |     | +   | +   | +   |     | +   | 7,77 | 9 |
|          | "2"  |    | +  | +  | +  |     | +   |     | +   | +   | +   | +   |     | +   | 5,06 | 9 |
| #3       | "1"  |    | +  | +  | +  | +   | +   | +   | +   | +   |     |     |     | +   | 7,65 | 9 |
|          | "2"  | +  | +  |    | +  |     |     | +   | +   |     | +   |     |     | +   | 6,39 | 7 |
| #4       | "1"  |    |    |    |    | +   |     |     | +   | +   |     | +   |     | +   | 4,8  | 5 |
|          | "2"  |    |    |    | +  |     |     | +   | +   |     |     | +   |     | +   | 2,94 | 5 |
| #5       | "1"  | +  | +  | +  |    | +   | +   |     |     |     | +   |     |     | +   | 4,94 | 7 |
|          | "2"  | +  | +  | +  |    | +   |     | +   |     |     | +   | +   | +   | +   | 5,22 | 9 |
| #6       | "1"  | +  | +  | +  | +  | +   | +   |     |     | +   |     |     |     | +   | 8    | 9 |
|          | "2"  | +  | +  | +  | +  |     | +   | +   |     | +   | +   |     |     | +   | 6,42 | 9 |
| #7       | "1"  | +  | +  | +  | +  |     | +   |     |     | +   |     |     | +   | +   | 5,24 | 8 |
|          | "2"  | +  | +  | +  | +  | +   |     |     | +   | +   |     | +   |     | +   | 5,01 | 9 |
| #8       | "1"  |    | +  | +  | +  |     |     | +   |     |     | +   |     |     | +   | 2,22 | 6 |
|          | "2"  | +  | +  | +  | +  | +   | +   |     |     |     | +   | +   |     | +   | 5,23 | 9 |
| #9       | "1"  |    |    | +  | +  |     | +   |     | +   |     |     | +   | +   | +   | 4,47 | 7 |
|          | "2"  |    | +  | +  | +  |     | +   |     | +   | +   |     | +   | +   | +   | 5,28 | 9 |
| #10      | "1"  |    |    |    |    |     | +   |     | +   | +   |     |     |     | +   | 2,03 | 4 |
|          | "2"  | +  |    | +  | +  | +   | +   |     |     |     | +   | +   | +   | +   | 6,02 | 9 |
| #11      | "1"  | +  | +  |    |    |     | +   |     | +   | +   |     |     |     | +   | 3,1  | 6 |
|          | "2"  | +  | +  | +  | +  |     |     | +   | +   |     | +   | +   |     | +   | 5,75 | 9 |
| #12      | "1"  | +  |    | +  |    |     |     | +   | +   | +   | +   | +   | +   | +   | 5,57 | 9 |
|          | "2"  |    |    | +  |    |     |     |     |     | +   |     |     |     | +   | 0,71 | 3 |
| #13      | "1"  |    | +  | +  |    | +   | +   |     |     |     |     | +   |     | +   | 5,39 | 6 |
|          | "2"  | +  | +  | +  | +  | +   |     | +   |     |     | +   | +   |     | +   | 5,05 | 9 |
| #14      | "1"  |    |    | +  |    |     |     | +   | +   |     |     | +   | +   | +   | 3,24 | 6 |
|          | "2"  | +  |    |    | +  |     | +   | +   |     | +   | +   |     | +   | +   | 5,93 | 8 |
| #15      | "1"  | +  | +  | +  | +  | +   | +   |     |     | +   |     | +   |     | +   | 5,39 | 9 |
|          | "2"  |    | +  | +  | +  | +   | +   | +   |     |     | +   | +   |     | +   | 5,13 | 9 |
| #16      | "1"  | +  |    |    | +  | +   | +   | +   | +   |     | +   | +   |     | +   | 5,01 | 9 |
|          | "2"  | +  |    |    |    | +   |     | +   | +   | +   | +   | +   | +   | +   | 5,22 | 9 |
| #17      | "1"  |    |    |    | +  |     |     |     |     | +   |     |     |     | +   | 2,96 | 3 |
|          | "2"  | +  | +  |    | +  |     | +   |     | +   | +   | +   |     | +   | +   | 5,12 | 9 |
| #18      | "1"  | +  | +  | +  | +  | +   |     |     | +   |     | +   |     | +   | +   | 5,29 | 9 |
|          | "2"  |    |    | +  |    | +   |     |     | +   | +   |     | +   |     | +   | 5,74 | 6 |
| #19      | "1"  | +  | +  | +  | +  |     | +   |     | +   |     | +   |     | +   | +   | 7,11 | 9 |
|          | "2"  | +  |    | +  | +  |     | +   | +   |     | +   |     | +   | +   | +   | 5,38 | 9 |
| #20      | "1"  |    | +  | +  | +  | +   | +   |     |     | +   |     | +   | +   | +   | 5,48 | 9 |
|          | "2"  |    |    |    | +  |     |     | +   |     |     | +   |     | +   | +   | 4,65 | 5 |
| #21      | "1"  |    |    |    |    |     | +   | +   |     | +   |     | +   |     | +   | 4,94 | 5 |
|          | "2"  |    | +  | +  | +  |     |     | +   | +   | +   | +   | +   |     | +   | 5,68 | 9 |
| #22      | "1"  | +  | +  |    | +  |     |     |     | +   |     |     |     | +   | +   | 3,75 | 6 |

|     |     |   |   |   |   |   |   |   |   |   |   |   |   |   |       |   |
|-----|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|-------|---|
|     | "2" | + |   | + | + | + |   | + | + |   | + | + |   | + | 5,14  | 9 |
| #23 | "1" | + | + | + | + | + | + |   | + | + |   |   |   | + | 5,25  | 9 |
|     | "2" | + |   | + |   |   | + | + | + |   | + | + | + | + | 5,12  | 9 |
| #24 | "1" | + | + | + | + |   |   | + | + |   |   | + | + | + | 6,71  | 9 |
|     | "2" | + | + |   | + |   |   | + |   |   |   | + |   | + | 2,34  | 6 |
| #25 | "1" |   | + | + |   | + | + | + | + |   | + | + |   | + | 5,04  | 9 |
|     | "2" | + | + | + | + |   |   | + | + |   | + |   | + | + | 5,05  | 9 |
| #26 | "1" |   | + |   | + | + |   |   | + | + | + | + |   | + | 6,84  | 8 |
|     | "2" |   | + | + |   | + |   |   |   | + |   |   | + | + | 3,72  | 6 |
| #27 | "1" |   |   | + | + | + | + |   | + |   |   |   |   | + | 3,47  | 6 |
|     | "2" |   | + | + | + |   | + | + | + |   | + | + |   | + | 5,28  | 9 |
| #28 | "1" | + |   | + | + | + | + |   | + |   | + |   | + | + | 8,41  | 9 |
|     | "2" | + | + | + |   |   |   |   |   |   |   | + |   | + | 1,55  | 5 |
| #29 | "1" | + | + | + | + | + |   | + |   |   |   | + | + | + | 12,15 | 9 |
|     | "2" | + | + | + |   | + | + | + |   |   |   |   | + | + | 7,52  | 8 |
| #30 | "1" | + |   | + | + | + | + |   | + |   | + | + |   | + | 5,27  | 9 |
|     | "2" |   |   |   | + |   |   |   | + |   |   | + |   | + | 1,18  | 4 |
| #31 | "1" | + | + |   | + |   | + |   | + | + | + | + |   | + | 8,81  | 9 |
|     | "2" |   | + | + | + |   | + | + | + |   | + |   |   | + | 5,7   | 8 |
| #32 | "1" | + | + |   | + | + | + |   |   | + |   | + | + | + | 8,38  | 9 |
|     | "2" | + | + | + |   |   | + |   | + | + | + | + |   | + | 5,21  | 9 |
| #33 | "1" |   | + |   | + |   |   | + |   |   | + |   | + | + | 4,49  | 6 |
|     | "2" | + |   |   | + | + |   |   | + | + | + | + | + | + | 5,71  | 9 |
| #34 | "1" | + |   | + | + | + |   |   | + | + |   | + | + | + | 5,39  | 9 |
|     | "2" | + |   | + | + | + |   |   |   | + | + | + | + | + | 5,24  | 9 |
| #35 | "1" | + | + |   | + |   |   |   |   |   |   |   | + | + | 4,16  | 5 |
|     | "2" | + |   |   | + |   |   |   |   |   | + | + |   | + | 2,62  | 5 |
| #36 | "1" | + | + | + | + | + |   | + |   |   |   | + | + | + | 5,2   | 9 |
|     | "2" | + |   | + |   | + | + | + | + |   | + | + |   | + | 5,1   | 9 |
| #37 | "1" |   | + | + |   | + |   | + | + |   |   | + | + | + | 7,43  | 8 |
|     | "2" | + | + |   |   | + |   | + |   | + | + | + |   | + | 7,58  | 8 |
| #38 | "1" | + |   |   |   |   |   |   | + |   | + |   | + | + | 4,38  | 5 |
|     | "2" | + | + |   | + |   | + |   | + |   | + | + | + | + | 5,38  | 9 |
| #39 | "1" |   | + | + | + | + | + | + |   |   |   | + | + | + | 5,13  | 9 |
|     | "2" | + | + | + |   | + |   |   |   |   | + |   |   | + | 5,21  | 6 |
| #40 | "1" |   |   | + |   |   | + |   |   |   | + |   | + | + | 4,67  | 5 |
|     | "2" |   | + | + | + |   |   |   |   | + |   |   |   | + | 2,17  | 5 |
| #41 | "1" | + | + |   |   | + | + |   | + | + | + |   | + | + | 7,48  | 9 |
|     | "2" | + |   | + | + | + |   |   | + | + |   | + | + | + | 5,11  | 9 |
| #42 | "1" |   |   |   |   |   | + | + |   |   | + | + |   | + | 2,5   | 5 |
|     | "2" | + |   | + |   | + |   |   | + |   |   |   |   | + | 2,73  | 5 |
| #43 | "1" | + | + | + | + | + | + | + |   |   |   | + |   | + | 5,04  | 9 |
|     | "2" | + |   | + |   |   | + |   |   |   | + |   | + | + | 4,5   | 6 |
| #44 | "1" |   | + |   | + |   |   |   | + | + |   | + | + | + | 3,68  | 7 |
|     | "2" | + |   |   |   | + | + |   | + | + |   | + |   | + | 6,06  | 7 |
| #45 | "1" |   | + | + |   | + |   | + | + | + | + | + |   | + | 5,35  | 9 |
|     | "2" |   | + | + | + | + |   | + |   | + | + | + |   | + | 5,09  | 9 |
| #46 | "1" | + |   | + | + |   |   | + |   |   |   |   | + | + | 4,72  | 6 |
|     | "2" | + | + | + | + |   |   | + |   | + | + | + |   | + | 5,16  | 9 |
| #47 | "1" | + | + |   |   |   | + |   |   | + |   |   |   | + | 1,84  | 5 |
|     | "2" | + | + | + | + | + |   |   |   | + |   | + | + | + | 5,14  | 9 |
| #48 | "1" |   | + | + | + | + | + |   |   |   | + | + | + | + | 5,18  | 9 |
|     | "2" | + | + |   | + | + | + | + |   |   | + | + |   | + | 5,36  | 9 |
| #49 | "1" |   | + | + |   | + |   |   |   |   | + |   | + | + | 3,11  | 5 |
|     | "2" | + | + | + | + | + |   |   | + |   | + |   | + | + | 5,67  | 9 |
| #50 | "1" |   | + | + | + | + | + |   | + |   | + | + |   | + | 7,57  | 9 |
|     | "2" |   |   | + |   |   |   |   | + |   | + | + |   | + | 2,65  | 5 |

|     |     |   |   |   |   |   |   |   |   |   |   |   |   |   |      |   |
|-----|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|------|---|
| #51 | "1" |   | + | + | + | + | + |   | + | + | + |   |   | + | 5,22 | 9 |
|     | "2" | + |   | + | + | + | + | + |   |   | + |   | + | + | 6,35 | 9 |
| #52 | "1" | + | + | + | + | + |   | + | + |   | + |   |   | + | 5,16 | 9 |
|     | "2" | + | + |   | + |   | + |   | + |   | + | + | + | + | 5,08 | 9 |
| #53 | "1" | + | + | + |   | + |   |   | + | + | + |   | + | + | 5,38 | 9 |
|     | "2" |   |   | + | + | + |   | + | + |   |   |   |   | + | 3,93 | 6 |
| #54 | "1" |   | + |   |   | + |   | + | + |   | + |   | + | + | 5,92 | 7 |
|     | "2" | + |   | + | + | + | + |   | + |   |   | + | + | + | 8,69 | 9 |
| #55 | "1" | + |   |   |   |   | + |   |   |   |   |   | + | + | 2,3  | 4 |
|     | "2" |   | + | + | + | + | + |   | + |   | + | + |   | + | 5,21 | 9 |
| #56 | "1" | + |   | + |   | + | + | + | + |   | + |   | + | + | 8,13 | 9 |
|     | "2" | + | + |   | + |   | + | + |   |   | + | + |   | + | 7,58 | 8 |
| #57 | "1" |   | + | + |   |   | + | + | + | + |   | + | + | + | 8,19 | 9 |
|     | "2" | + |   | + |   |   | + |   | + | + | + | + | + | + | 5,73 | 9 |
| #58 | "1" |   | + | + |   | + | + | + |   | + | + |   | + | + | 5,98 | 9 |
|     | "2" | + | + |   | + |   |   | + | + | + | + |   | + | + | 8,96 | 9 |
| #59 | "1" | + |   | + | + | + | + |   |   | + |   | + | + | + | 6,46 | 9 |
|     | "2" |   | + | + |   | + |   | + |   | + | + | + | + | + | 5,44 | 9 |
| #60 | "1" |   | + | + |   |   | + | + |   |   |   | + |   | + | 3,93 | 6 |
|     | "2" | + | + |   | + | + |   | + |   |   | + |   |   | + | 6,03 | 7 |
| #61 | "1" |   | + | + | + | + |   | + |   |   |   | + | + | + | 7,52 | 8 |
|     | "2" |   |   | + |   | + |   |   | + | + |   |   |   | + | 3,79 | 5 |
| #62 | "1" | + | + |   |   | + | + | + |   | + |   | + | + | + | 8,31 | 9 |
|     | "2" |   | + |   | + | + | + |   | + | + | + |   | + | + | 5,45 | 9 |
| #63 | "1" |   | + | + | + |   |   |   |   | + |   |   | + | + | 3,56 | 6 |
|     | "2" |   |   | + |   |   |   |   |   |   |   | + |   | + | 0,69 | 3 |
| #64 | "1" |   |   |   | + |   |   | + |   |   | + |   | + | + | 4,72 | 5 |
|     | "2" | + |   | + | + |   | + | + | + |   | + | + |   | + | 5,03 | 9 |
| #65 | "1" | + |   |   | + | + | + | + | + |   |   | + | + | + | 5,1  | 9 |
|     | "2" |   | + | + |   |   | + |   | + |   |   | + | + | + | 6,89 | 7 |
| #66 | "1" | + | + | + | + | + | + |   |   | + | + |   |   | + | 8,34 | 9 |
|     | "2" |   | + | + | + |   |   | + | + | + | + |   | + | + | 5,22 | 9 |
| #67 | "1" |   | + |   | + |   | + |   | + | + |   |   | + | + | 5,67 | 7 |
|     | "2" | + |   |   | + |   | + | + |   |   |   | + |   | + | 3,61 | 6 |
| #68 | "1" |   | + |   | + |   |   |   | + |   |   |   | + | + | 4,38 | 5 |
|     | "2" |   |   |   | + |   | + |   |   |   |   | + | + | + | 4,95 | 5 |
| #69 | "1" | + |   | + |   | + |   | + | + |   | + |   | + | + | 5,74 | 8 |
|     | "2" |   | + | + | + | + |   | + | + | + | + |   |   | + | 5,27 | 9 |
| #70 | "1" | + | + | + | + | + | + |   | + |   |   | + |   | + | 8,81 | 9 |
|     | "2" |   | + |   | + | + | + | + | + |   | + |   | + | + | 5,1  | 9 |
| #71 | "1" |   |   |   |   |   |   | + |   |   |   |   | + | + | 2,76 | 3 |
|     | "2" |   |   | + |   | + |   | + |   |   |   |   | + | + | 3,5  | 5 |
| #72 | "1" |   | + |   |   |   |   |   |   |   | + |   |   | + | 2,6  | 3 |
|     | "2" |   | + |   | + |   |   | + | + |   |   |   |   | + | 4,66 | 5 |
| #73 | "1" | + | + |   |   |   |   |   |   |   |   | + |   | + | 1,28 | 4 |
|     | "2" |   | + | + |   | + | + |   | + | + |   | + |   | + | 5,89 | 8 |
| #74 | "1" | + |   |   | + | + | + | + |   | + | + |   | + | + | 5,09 | 9 |
|     | "2" | + |   | + |   | + |   |   |   | + |   |   |   | + | 3,68 | 5 |
| #75 | "1" | + | + |   |   | + | + | + | + |   |   |   |   | + | 4,29 | 7 |
|     | "2" | + |   |   | + |   |   |   | + | + |   |   |   | + | 4,15 | 5 |
| #76 | "1" | + | + | + | + |   |   | + | + |   |   | + | + | + | 8,02 | 9 |
|     | "2" |   |   | + |   | + |   | + |   |   |   | + |   | + | 4,36 | 5 |
| #77 | "1" |   |   | + |   | + |   | + | + |   |   |   |   | + | 2,67 | 5 |
|     | "2" |   | + | + |   | + | + |   | + | + | + |   | + | + | 5,02 | 9 |
| #78 | "1" | + | + |   |   |   | + |   | + | + | + | + | + | + | 6,63 | 9 |
|     | "2" |   |   |   |   | + |   |   | + |   |   |   | + | + | 1,1  | 4 |
| #79 | "1" | + |   | + |   | + | + | + |   |   | + | + | + | + | 8,06 | 9 |
|     | "2" | + | + | + | + |   | + |   |   | + |   | + | + | + | 8,41 | 9 |

|     |     |   |   |   |   |   |   |   |   |   |   |   |   |      |   |
|-----|-----|---|---|---|---|---|---|---|---|---|---|---|---|------|---|
| #80 | "1" |   |   | + |   |   |   |   | + |   |   | + | + | 3,34 | 4 |
|     | "2" |   |   | + | + | + |   |   | + | + |   |   | + | 5,78 | 6 |
| #81 | "1" | + |   |   |   | + |   |   | + | + |   | + | + | 4,26 | 6 |
|     | "2" |   |   | + |   |   |   |   | + | + | + | + | + | 4,28 | 6 |
| #82 | "1" | + |   |   |   | + |   | + |   |   |   | + | + | 3,66 | 5 |
|     | "2" |   |   |   | + |   |   | + |   |   | + |   | + | 2,89 | 4 |
| #83 | "1" | + | + |   | + | + |   | + | + | + | + |   | + | 5,33 | 9 |
|     | "2" | + |   | + | + |   | + | + |   | + | + |   | + | 5,26 | 9 |
| #84 | "1" |   |   | + |   |   |   |   | + | + | + |   | + | 4,44 | 6 |
|     | "2" |   | + |   |   |   | + |   | + | + | + | + | + | 6,81 | 8 |
| #85 | "1" |   |   |   |   |   | + |   | + | + |   | + | + | 3,89 | 5 |
|     | "2" | + |   | + |   |   |   |   | + | + |   |   | + | 4,23 | 5 |
| #86 | "1" |   |   |   |   |   | + |   |   |   |   | + | + | 1,01 | 3 |
|     | "2" |   | + | + | + | + |   | + |   | + | + | + | + | 5,68 | 9 |
| #87 | "1" |   | + | + | + |   | + |   | + |   |   | + | + | 6,64 | 7 |
|     | "2" | + | + |   | + | + |   | + | + |   |   | + | + | 5,96 | 9 |
| #88 | "1" |   | + |   |   |   |   |   | + | + |   |   | + | 3,36 | 4 |
|     | "2" |   |   | + | + | + |   |   | + |   |   | + | + | 5,26 | 6 |
| #89 | "1" | + |   |   | + |   |   |   | + |   |   | + | + | 3,65 | 5 |
|     | "2" | + |   | + |   |   | + | + | + | + |   | + | + | 5,61 | 9 |

В Таблица 4.8 и за двата варианта са посочени брой на редуцираните признаци и съответния брой участници.

Таблица 4.8 Брой срещания на редуцирани признаци

| Вар. | Брой на редуцирани признаци | Брой участници | Вар. | Брой на редуцирани признаци | Брой участници |
|------|-----------------------------|----------------|------|-----------------------------|----------------|
| "1"  | 9                           | 42             | "2"  | 9                           | 48             |
| "1"  | 8                           | 5              | "2"  | 8                           | 9              |
| "1"  | 7                           | 7              | "2"  | 7                           | 4              |
| "1"  | 6                           | 12             | "2"  | 6                           | 10             |
| "1"  | 5                           | 14             | "2"  | 5                           | 15             |
| "1"  | 4                           | 5              | "2"  | 4                           | 3              |
| "1"  | 3                           | 4              | "2"  | 3                           | 0              |

Вижда се, че е постигнато значително редуциране на броя на признаците, понеже първоначалния им брой (равен на 13) се редуцира до 9 за около половината участници и при двата варианта, а за 30% от участниците (26 души) и при двата варианта се намалява до 5 или 6 признака. При малък брой участници признаците достигат брой, по-малък от 5 (за Вариант "1" – 9 души, а за Вариант "2" – само 3). Тези резултати показват, че първоначалният брой признаци се редуцира значително и е по-малък при използване на неумели фалшификати за отрицателни примери (Вариант "1").

Таблица 4.9 съдържа обобщена информация за всеки от построените модели (с номера от 1 до 6) за базата подписи *SU*sig. Те се различават по използваните отрицателни примери за обучение и по броя на признаците.

За всеки участник се извършва оценка на точността за всяка НМ по метода на 10-кратната крос валидация и оценка на точността на всеки класификатори *k*NN по LOOCV, а за окончателен се избира този модел и съответната стойност на *H* или *k*, който демонстрира най-висока точност. В Таблица 4.10 са посочени параметрите на така избраните модели за всеки участник.

Таблица 4.9 Параметри на моделите за базата SUsig

| # на модел | Брой признаци (входни неврони) | Брой собствени подписи | Брой подправени подписи |            | Брой скрити неврони $H$ | Брой съседни $k$ |
|------------|--------------------------------|------------------------|-------------------------|------------|-------------------------|------------------|
| 1          | Общи признаци                  | 8 или 10               | 15 неумели              | Случай “1” | от 1 до 5               | 1 или 3          |
| 2          | Вариант “1”                    |                        |                         |            |                         |                  |
| 3          | Вариант “2”                    |                        |                         |            |                         |                  |
| 4          | Вариант “2”                    |                        | 9 неумели и 6 умели     | Случай “2” |                         |                  |
| 5          | Вариант “1”                    |                        |                         |            |                         |                  |
| 6          | Общи признаци                  |                        |                         |            |                         |                  |

Таблица 4.10 Окончателни модели за всеки участник от SUsig

| Участник | № на избран модел за НМ | Брой скрити неврони $H$ | Точност по 10-кратна CV (%) | № на избран модел за kNN | Брой съседни $k$ | Точност по LOOCV (%) |
|----------|-------------------------|-------------------------|-----------------------------|--------------------------|------------------|----------------------|
| #1       | 1                       | 5                       | 100                         | 2                        | 1                | 100.00               |
| #2       | 3                       | 3                       | 98.33                       | 4                        | 1                | 95.65                |
| #3       | 5                       | 5                       | 93.7                        | 5                        | 3                | 86.96                |
| #4       | 4                       | 2                       | 95.93                       | 3                        | 1                | 92.00                |
| #5       | 2                       | 5                       | 98.89                       | 2                        | 1                | 96.00                |
| #6       | 4                       | 4                       | 98.15                       | 5                        | 1                | 100.00               |
| #7       | 3                       | 4                       | 97.41                       | 5                        | 1                | 95.65                |
| #8       | 1                       | 3                       | 97.41                       | 4                        | 3                | 86.96                |
| #9       | 5                       | 2                       | 98.89                       | 4                        | 1                | 96.00                |
| #10      | 4                       | 4                       | 99.07                       | 2                        | 3                | 96.00                |
| #11      | 6                       | 2                       | 98.52                       | 3                        | 1                | 92.00                |
| #12      | 6                       | 1                       | 99.63                       | 5                        | 1                | 92.00                |
| #13      | 1                       | 1                       | 93.33                       | 5                        | 1                | 96.00                |
| #14      | 2                       | 3                       | 99.26                       | 2                        | 3                | 100.00               |
| #15      | 2                       | 4                       | 99.26                       | 2                        | 3                | 100.00               |
| #16      | 2                       | 2                       | 99.07                       | 2                        | 1                | 100.00               |
| #17      | 3                       | 1                       | 97.78                       | 3                        | 1                | 100.00               |
| #18      | 2                       | 1                       | 96.48                       | 2                        | 1                | 92.00                |
| #19      | 5                       | 4                       | 99.63                       | 5                        | 1                | 100.00               |
| #20      | 6                       | 3                       | 97.78                       | 5                        | 1                | 95.65                |
| #21      | 1                       | 5                       | 99.26                       | 1                        | 1                | 100.00               |
| #22      | 5                       | 5                       | 97.41                       | 2                        | 1                | 95.65                |
| #23      | 1                       | 3                       | 90                          | 4                        | 1                | 80.00                |
| #24      | 5                       | 3                       | 98.7                        | 2                        | 1                | 96.00                |
| #25      | 2                       | 2                       | 97.78                       | 4                        | 1                | 100.00               |
| #26      | 4                       | 2                       | 97.22                       | 4                        | 1                | 100.00               |
| #27      | 2                       | 4                       | 97.78                       | 5                        | 1                | 96.00                |
| #28      | 2                       | 2                       | 98.89                       | 4                        | 1                | 96.00                |
| #29      | 2                       | 4                       | 99.63                       | 4                        | 1                | 96.00                |
| #30      | 3                       | 3                       | 97.04                       | 4                        | 1                | 92.00                |

|     |   |   |       |   |   |        |
|-----|---|---|-------|---|---|--------|
| #31 | 6 | 3 | 99.44 | 4 | 1 | 100.00 |
| #32 | 2 | 2 | 98.15 | 5 | 1 | 100.00 |
| #33 | 4 | 4 | 98.15 | 2 | 1 | 91.30  |
| #34 | 2 | 2 | 97.96 | 2 | 1 | 91.30  |
| #35 | 6 | 4 | 97.04 | 4 | 1 | 91.30  |
| #36 | 2 | 4 | 99.44 | 2 | 1 | 95.65  |
| #37 | 2 | 5 | 98.33 | 5 | 1 | 92.00  |
| #38 | 1 | 2 | 99.63 | 2 | 1 | 100.00 |
| #39 | 4 | 3 | 98.52 | 1 | 1 | 100.00 |
| #40 | 2 | 2 | 94.07 | 5 | 1 | 88.00  |
| #41 | 1 | 5 | 98.89 | 3 | 1 | 100.00 |
| #42 | 2 | 4 | 98.52 | 2 | 1 | 96.00  |
| #43 | 4 | 2 | 99.26 | 3 | 1 | 96.00  |
| #44 | 4 | 2 | 97.59 | 4 | 1 | 100.00 |
| #45 | 3 | 5 | 96.3  | 5 | 1 | 92.00  |
| #46 | 1 | 5 | 96.67 | 2 | 3 | 92.00  |
| #47 | 3 | 2 | 99.26 | 2 | 1 | 100.00 |
| #48 | 4 | 4 | 97.41 | 4 | 1 | 86.96  |
| #49 | 2 | 2 | 100   | 3 | 1 | 95.65  |
| #50 | 1 | 3 | 97.04 | 3 | 1 | 96.00  |
| #51 | 1 | 3 | 97.96 | 2 | 1 | 96.00  |
| #52 | 2 | 3 | 99.44 | 1 | 1 | 100.00 |
| #53 | 3 | 4 | 97.59 | 2 | 3 | 96.00  |
| #54 | 3 | 4 | 96.11 | 2 | 1 | 92.00  |
| #55 | 5 | 3 | 99.07 | 5 | 1 | 100.00 |
| #56 | 5 | 4 | 98.52 | 5 | 1 | 92.00  |
| #57 | 6 | 5 | 99.44 | 3 | 1 | 100.00 |
| #58 | 6 | 4 | 91.67 | 2 | 1 | 76.00  |
| #59 | 2 | 4 | 99.63 | 5 | 1 | 100.00 |
| #60 | 6 | 2 | 96.3  | 5 | 3 | 88.00  |
| #61 | 2 | 4 | 98.33 | 5 | 1 | 100.00 |
| #62 | 6 | 2 | 97.59 | 4 | 1 | 96.00  |
| #63 | 6 | 2 | 99.44 | 5 | 1 | 100.00 |
| #64 | 3 | 5 | 99.07 | 2 | 1 | 100.00 |
| #65 | 2 | 5 | 98.52 | 3 | 1 | 100.00 |
| #66 | 1 | 2 | 99.26 | 3 | 1 | 100.00 |
| #67 | 3 | 3 | 99.26 | 2 | 1 | 100.00 |
| #68 | 2 | 2 | 98.7  | 3 | 1 | 96.00  |
| #69 | 3 | 4 | 98.33 | 3 | 1 | 100.00 |
| #70 | 5 | 2 | 100   | 4 | 3 | 95.65  |
| #71 | 3 | 5 | 95.56 | 4 | 1 | 96.00  |
| #72 | 1 | 2 | 99.44 | 2 | 1 | 100.00 |
| #73 | 2 | 4 | 100   | 2 | 1 | 100.00 |
| #74 | 4 | 4 | 99.63 | 4 | 1 | 100.00 |
| #75 | 2 | 5 | 97.22 | 2 | 1 | 96.00  |
| #76 | 4 | 5 | 92.78 | 6 | 3 | 84.00  |
| #77 | 5 | 3 | 95.93 | 5 | 1 | 95.65  |

|     |   |   |       |   |   |        |
|-----|---|---|-------|---|---|--------|
| #78 | 2 | 5 | 98.52 | 3 | 1 | 100.00 |
| #79 | 2 | 4 | 95.19 | 2 | 1 | 92.00  |
| #80 | 4 | 2 | 97.96 | 4 | 1 | 96.00  |
| #81 | 6 | 5 | 97.96 | 5 | 1 | 100.00 |
| #82 | 4 | 4 | 99.63 | 4 | 1 | 100.00 |
| #83 | 3 | 4 | 100   | 3 | 1 | 100.00 |
| #84 | 2 | 1 | 98.89 | 3 | 1 | 100.00 |
| #85 | 2 | 2 | 97.78 | 2 | 1 | 100.00 |
| #86 | 1 | 5 | 96.85 | 4 | 1 | 100.00 |
| #87 | 4 | 1 | 96.48 | 1 | 1 | 100.00 |
| #88 | 4 | 3 | 100   | 1 | 1 | 100.00 |
| #89 | 4 | 3 | 99.63 | 4 | 1 | 100.00 |

### Сравнение на точността на класификаторите НМ и $k$ NN

Целта на сравнението е да се провери как избраният подход за построяване на класификатор (НМ) се съотнася към други широко използвани класификатори. При малки извадки, когато оптимални в статистически смисъл подходи, не са приложими, най-често се използват евристични подходи като  $k$  – най-близки съседи. Следва да се отбележи, че при голям брой наблюдения и голямо  $k$  този подход е квази оптимален, т.е. води до получаване на класификатор, близък до оптималния Бейсов класификатор. В Таблици 4.10 – 4.12 са дадени резултатите от проведеното изследване.

Средната стойност на точността на НМ, оценена за всички участници от SUsig по избраните за тях модели чрез 10-кратна крос-валидация, е равна на 97.95%. Средната стойност на точността на  $k$ NN, оценена за всички участници от SUsig по избраните за тях модели чрез LOOCV, е равна на 96.13%. Тези резултати показват предимство за невронната мрежа. За отбелязване е, че най-добри резултати за  $k$ NN се получават при  $k = 1$  за 79 потребители или 89% от случаите. За изчислените средни стойности и средноквадратични отклонения е изчислена стойността на  $t$ -статистиката  $t = 3.29$ . При 176 степени на свобода тази стойност е значима с вероятност 0.99.

### Зависимост на точността на верификация от подмножеството признаци за разпознаване и вида фалшиви подписи за обучение

Изследването има за цел (1) да се оцени ефекта от редуцирането на броя на признаците върху точността на класификацията (2) да се изследва зависимостта на точността на класификация от вида фалшиви подписи, използвани за обучение на класификаторите. За целта изчисляваме средните стойности на оценената точност по най-добрите модели за НМ (Таблица 4.11) и  $k$ NN (Таблица 4.12), средната точност по Общи признаци, по признаци от Вариант “1” и по признаци от Вариант “2”.

Данните от Таблица 4.11 показват по-голяма точност на НМ при използване на индивидуалните признаци (Вариант “1” и Вариант “2”) спрямо точността по общи признаци като се наблюдава превес на точността по признаци от Вариант “1”. Точността по Вариант “1”, Случай “1” (неумели фалшификати за обучение) надвишава точността по Вариант “2”, Случай “2” (умели и неумели фалшификати за обучение).

Таблица 4.11 Таблица със средни стойности на оценената точност по най-добрите модели на НМ

|                | Общи признаци                           | Вариант 1                               | Вариант 2                               | Средна точност | Брой случаи |
|----------------|-----------------------------------------|-----------------------------------------|-----------------------------------------|----------------|-------------|
| Случай 1       | <b>97.36</b><br>(Модел 1,<br>13 случая) | <b>98.36</b><br>(Модел 2,<br>27 случая) | <b>97.85</b><br>(Модел 3,<br>13 случая) | <b>97.99</b>   | <b>53</b>   |
| Случай 2       | <b>97.71</b><br>(Модел 6,<br>11 случая) | <b>97.98</b><br>(Модел 5,<br>9 случая)  | <b>97.96</b><br>(Модел 4,<br>16 случая) | <b>97.89</b>   | <b>36</b>   |
| Средна точност | <b>97.52</b>                            | <b>98.27</b>                            | <b>97.91</b>                            | -              | -           |
| Брой случаи    | <b>24</b>                               | <b>36</b>                               | <b>29</b>                               | -              | <b>89</b>   |

При прилагане на класификатор  $kNN$  Таблица (4.12) точността по общите признаци надвишава тази по индивидуалните. Точността по Вариант “2”, Случай “1” (неумели фалшификати за обучение) надвишава точността по Вариант “2”, Случай “2” (умели и неумели фалшификати за обучение).

Таблица 4.12 Таблица със средни стойности на оценената точност по най-добрите модели на  $kNN$ 

|                | Общи признаци                        | Вариант 1                               | Вариант 2                               | Средна точност | Брой случаи |
|----------------|--------------------------------------|-----------------------------------------|-----------------------------------------|----------------|-------------|
| Случай 1       | <b>100</b><br>(Модел 1,<br>5 случая) | <b>95.92</b><br>(Модел 2,<br>27 случая) | <b>97.84</b><br>(Модел 3,<br>15 случая) | <b>96.97</b>   | <b>47</b>   |
| Случай 2       | <b>84</b><br>(Модел 6,<br>1 случая)  | <b>95.5</b><br>(Модел 5,<br>20 случая)  | <b>95.45</b><br>(Модел 4,<br>21 случая) | <b>95.20</b>   | <b>42</b>   |
| Средна точност | <b>97.33</b>                         | <b>95.74</b>                            | <b>96.45</b>                            | -              | -           |
| Брой случаи    | <b>6</b>                             | <b>47</b>                               | <b>36</b>                               | -              | <b>89</b>   |

Сравнението на средната точност между НМ и  $kNN$  по общи и индивидуални признаци показва превъзходство на НМ.

Броят на избраните модели, обучени с неумели фалшификати (Случай “1”), се среща при повече от половината участници. Той е равен на 53 (при 60 % от участниците) за НМ и равен на 47 (при 53 % от участниците) за  $kNN$ . Броят на избраните модели, построени по признаци от Вариант “1” е най-голям за НМ (за 36 участника), както и за  $kNN$  (за 47 участника).

### Построяване на класификатор НМ и $kNN$ по избраните модели за всеки потребител

В Таблица 4.13 са показани резултатите от оценката на горните класификатори върху едни и същи извадки от 830 собствени и 890 фалшиви подписи. Около 70% от подписите се използват за проектиране на класификатора, а останалите 30% за тестване. При невронните мрежи за обучение са използвани по 3 или 4 собствени



подписа и 7 фалшиви, а валидацията е извършена върху 2 или 3 собствени и 3 фалшиви подписа за всеки потребител. Тестването е извършено върху 3 собствени и 5 фалшиви подписа. При 9-кратна инициализация е избрана тази невронна мрежа, за която е получена най-добра оценка. За  $kNN$  са използвани 5 или 7 собствени подписа и 10 фалшиви, а за тестване – 3 собствени и 5 фалшиви. Разликата в броя използвани собствени подписи за обучение се дължи на различния брой събрани подписи от всеки участник. В Таблица 4.13 за съответните модели за всеки участник са посочени стойностите на праг на НМ, номера на инициализация на теглата, стойности на FAR, FRR и точността по НМ и  $kNN$ .

Таблица 4.13 Оценка на качеството на класификация с НМ и  $kNN$  за всички участници

| Участник | Праг (НМ) | № на инициализация | FAR (НМ) | FRR (НМ) | Точност по НМ | FAR ( $kNN$ ) | FRR ( $kNN$ ) | Точност по $kNN$ |
|----------|-----------|--------------------|----------|----------|---------------|---------------|---------------|------------------|
| #1       | 0.82      | 7                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #2       | 0.82      | 6                  | 0        | 0        | 100           | 20            | 0             | 87.5             |
| #3       | 0.77      | 4                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #4       | 0.83      | 5                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #5       | 0.92      | 3                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #6       | 0.92      | 2                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #7       | 0.82      | 2                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #8       | 0.81      | 4                  | 0        | 0        | 100           | 20            | 33.33         | 75               |
| #9       | 0.92      | 1                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #10      | 0.92      | 7                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #11      | 0.91      | 3                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #12      | 0.87      | 4                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #13      | 0.92      | 2                  | 0        | 0        | 100           | 0             | 66.67         | 75               |
| #14      | 0.82      | 8                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #15      | 0.81      | 9                  | 0        | 0        | 100           | 40            | 0             | 75               |
| #16      | 0.82      | 7                  | 0        | 0        | 100           | 0             | 33.33         | 87.5             |
| #17      | 0.77      | 4                  | 0        | 0        | 100           | 0             | 33.33         | 87.5             |
| #18      | 0.92      | 6                  | 0        | 0        | 100           | 0             | 33.33         | 87.5             |
| #19      | 0.82      | 9                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #20      | 0.8       | 9                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #21      | 0.82      | 9                  | 0        | 0        | 100           | 20            | 0             | 87.5             |
| #22      | 0.86      | 2                  | 40       | 0        | 75            | 0             | 33.33         | 87.5             |
| #23      | 0.9       | 1                  | 0        | 0        | 100           | 40            | 0             | 75               |
| #24      | 0.92      | 2                  | 0        | 0        | 100           | 0             | 33.33         | 87.5             |
| #25      | 0.92      | 2                  | 0        | 0        | 100           | 20            | 66.67         | 62.5             |
| #26      | 0.92      | 3                  | 0        | 0        | 100           | 40            | 0             | 75               |
| #27      | 0.91      | 7                  | 0        | 0        | 100           | 20            | 33.33         | 75               |
| #28      | 0.92      | 1                  | 0        | 0        | 100           | 0             | 33.33         | 87.5             |
| #29      | 0.91      | 3                  | 0        | 0        | 100           | 20            | 0             | 87.5             |
| #30      | 0.82      | 1                  | 40       | 0        | 75            | 60            | 66.67         | 37.5             |
| #31      | 0.92      | 7                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #32      | 0.8       | 6                  | 0        | 0        | 100           | 0             | 0             | 100              |
| #33      | 0.8       | 1                  | 0        | 0        | 100           | 40            | 33.33         | 62.5             |
| #34      | 0.82      | 8                  | 0        | 0        | 100           | 20            | 0             | 87.5             |

|     |      |   |    |   |     |    |       |      |
|-----|------|---|----|---|-----|----|-------|------|
| #35 | 0.82 | 6 | 0  | 0 | 100 | 0  | 33.33 | 87.5 |
| #36 | 0.82 | 5 | 0  | 0 | 100 | 0  | 0     | 100  |
| #37 | 0.85 | 5 | 40 | 0 | 75  | 0  | 66.67 | 75   |
| #38 | 0.82 | 4 | 0  | 0 | 100 | 0  | 0     | 100  |
| #39 | 0.92 | 4 | 0  | 0 | 100 | 0  | 0     | 100  |
| #40 | 0.8  | 2 | 0  | 0 | 100 | 0  | 66.67 | 75   |
| #41 | 0.92 | 8 | 0  | 0 | 100 | 0  | 0     | 100  |
| #42 | 0.91 | 8 | 0  | 0 | 100 | 0  | 0     | 100  |
| #43 | 0.92 | 9 | 0  | 0 | 100 | 0  | 33.33 | 87.5 |
| #44 | 0.91 | 9 | 0  | 0 | 100 | 20 | 0     | 87.5 |
| #45 | 0.87 | 8 | 0  | 0 | 100 | 0  | 33.33 | 87.5 |
| #46 | 0.91 | 8 | 0  | 0 | 100 | 0  | 0     | 100  |
| #47 | 0.92 | 4 | 0  | 0 | 100 | 0  | 0     | 100  |
| #48 | 0.82 | 1 | 0  | 0 | 100 | 40 | 33.33 | 62.5 |
| #49 | 0.81 | 1 | 0  | 0 | 100 | 0  | 0     | 100  |
| #50 | 0.89 | 9 | 0  | 0 | 100 | 0  | 0     | 100  |
| #51 | 0.91 | 1 | 0  | 0 | 100 | 20 | 0     | 87.5 |
| #52 | 0.82 | 3 | 0  | 0 | 100 | 0  | 0     | 100  |
| #53 | 0.91 | 2 | 0  | 0 | 100 | 0  | 33.33 | 87.5 |
| #54 | 0.83 | 2 | 0  | 0 | 100 | 0  | 33.33 | 87.5 |
| #55 | 0.82 | 3 | 0  | 0 | 100 | 0  | 33.33 | 87.5 |
| #56 | 0.88 | 5 | 0  | 0 | 100 | 0  | 33.33 | 87.5 |
| #57 | 0.92 | 3 | 0  | 0 | 100 | 0  | 0     | 100  |
| #58 | 0.89 | 5 | 40 | 0 | 75  | 40 | 66.67 | 50   |
| #59 | 0.92 | 4 | 0  | 0 | 100 | 0  | 66.67 | 75   |
| #60 | 0.87 | 2 | 0  | 0 | 100 | 40 | 0     | 75   |
| #61 | 0.91 | 7 | 0  | 0 | 100 | 0  | 33.33 | 87.5 |
| #62 | 0.92 | 1 | 0  | 0 | 100 | 20 | 33.33 | 75   |
| #63 | 0.9  | 5 | 0  | 0 | 100 | 0  | 0     | 100  |
| #64 | 0.81 | 4 | 0  | 0 | 100 | 0  | 0     | 100  |
| #65 | 0.85 | 5 | 0  | 0 | 100 | 0  | 0     | 100  |
| #66 | 0.92 | 3 | 0  | 0 | 100 | 60 | 0     | 62.5 |
| #67 | 0.91 | 7 | 0  | 0 | 100 | 0  | 0     | 100  |
| #68 | 0.91 | 8 | 0  | 0 | 100 | 0  | 33.33 | 87.5 |
| #69 | 0.82 | 9 | 0  | 0 | 100 | 0  | 0     | 100  |
| #70 | 0.82 | 2 | 0  | 0 | 100 | 0  | 0     | 100  |
| #71 | 0.75 | 5 | 0  | 0 | 100 | 0  | 0     | 100  |
| #72 | 0.92 | 1 | 0  | 0 | 100 | 0  | 0     | 100  |
| #73 | 0.92 | 3 | 0  | 0 | 100 | 20 | 33.33 | 75   |
| #74 | 0.92 | 1 | 0  | 0 | 100 | 0  | 33.33 | 87.5 |
| #75 | 0.87 | 4 | 0  | 0 | 100 | 0  | 0     | 100  |
| #76 | 0.9  | 6 | 0  | 0 | 100 | 0  | 0     | 100  |
| #77 | 0.82 | 2 | 0  | 0 | 100 | 0  | 0     | 100  |
| #78 | 0.88 | 9 | 0  | 0 | 100 | 0  | 0     | 100  |
| #79 | 0.86 | 9 | 0  | 0 | 100 | 20 | 0     | 87.5 |
| #80 | 0.65 | 3 | 40 | 0 | 75  | 20 | 33.33 | 75   |

|     |      |   |    |   |      |    |       |      |
|-----|------|---|----|---|------|----|-------|------|
| #81 | 0.92 | 4 | 0  | 0 | 100  | 0  | 33.33 | 87.5 |
| #82 | 0.82 | 8 | 0  | 0 | 100  | 0  | 33.33 | 87.5 |
| #83 | 0.81 | 5 | 0  | 0 | 100  | 0  | 0     | 100  |
| #84 | 0.9  | 8 | 0  | 0 | 100  | 0  | 0     | 100  |
| #85 | 0.9  | 6 | 0  | 0 | 100  | 0  | 0     | 100  |
| #86 | 0.92 | 7 | 0  | 0 | 100  | 20 | 0     | 87.5 |
| #87 | 0.78 | 2 | 0  | 0 | 100  | 0  | 0     | 100  |
| #88 | 0.92 | 9 | 0  | 0 | 100  | 20 | 0     | 87.5 |
| #89 | 0.82 | 3 | 40 | 0 | 87.5 | 20 | 0     | 87.5 |

Получените стойности за средната точност, лъжливо положителните и лъжливо отрицателните отговори за невронната мрежа са съответно 98.46 %, 2.70 % и 0 %, а за kNN – 89.47 %, 8.09 % и 14.61 %. Те показват превъзходство на невронната мрежа спрямо kNN класификатора. Изчислената стойност на t-статистиката  $t = 5.98$  показва значима разлика с вероятност 0.99.

### Сравнение на резултатите с тези от други автори

Върху същата база данни за подписи SUsig е проектиран класификатор от авторите *Yanikoglu* и *Kholmatov* [3], който е оценен с 1.64% *FRR* и 1.28% *FAR*.

В [13] са тествани различни архитектури на невронни мрежи за верификация на подпис. Най-добрият получен експериментален резултат е 2% *FAR* и 1.3% *FRR* за многослоен перцептрон с един скрит слой. Във всичките експерименти са използвани по 5 подписа от всеки потребител за обучение и 41 глобални признаци на входа.

В [37] е представен класификатор kNN за оф-лайн подписи. Стойността на точността е оценена чрез крос валидация и е равен на 97.47%. В [38] и [39] са представени класификатори kNN за оф-лайн подписи, за които е постигната точност съответно от 70% и 78%.

Така представените в дисертацията класификатори имат оценки сходни с тези, които се срещат в литературата с използване на същата база данни за подписи, както и със същите по вид класификатори.

## 4.5 Резултати в Глава 4

От проведените експерименти за избор на най-информативно подмножество от признаци от собствената база от подписи и базата SUsig, както и за точността на верификацията, могат да се направят следните изводи:

1. Последователното използване на метода на корелационните плеади и регресионния подход, основан на *Cr* критерия на Mallows, води до съществено намаляване на броя на необходимите признаци (приблизително два пъти).
2. За всеки от участниците в експериментите съществува специфично подмножество от признаци, което описва най-добре съвкупността от неговите подписи или с други думи, не съществува подмножество от признаци, което еднакво добре да разграничава всеки от участниците. За някои от участниците индивидуалното подмножество се състои от 3-5 признака.
3. По-значителна редукция в броя на признаците се получава при използване на неумели фалшификати (Вариант “1”) като отрицателни примери.
4. Редукцията на броя на признаците при невронните мрежи води до повишаване на точността на верификация, докато при kNN се наблюдава намаляване на точността.

5. Обучените с неумели фалшификати класификатори (Случай "1") водят до по-добри резултати в сравнение с обучените с умели и неумели фалшификати (Случай "2").

Тези изводи водят до следните препоръки за прилагане на изложения в дисертационната работа комбиниран метод за верификация по подпис:

- 1) да се редуцира броя на признаците с последователно прилагане на метода на корелационните плеади и  $C_p$  критерия на Mallows;
- 2) Да се обучат класификатори по НМ с използване на неумели фалшификати и да се избере най-точния от тях за всеки конкретен участник.

## Глава 5. Софтуерно приложение

Пета глава включва следните раздели.

### 5.1 Бази от данни за подписи

Експериментите с класификатори за верификация по подпис са извършени върху базата данни за подписи, събрана за целта, както и върху публично достъпната база данни за подписи SUsig.

#### 5.1.1 Събиране на подписи

Подписите се събират мобилно с помощта на лаптоп и графичен таблет *Wacom Intuos3 A5 PTZ-630* [61], който дава възможност за визуална обратна връзка и позволява писането както със стандартна писалка за таблет (*grip pen*), така и с писалка с мастило (*inking pen*), с която може да се пише по стандартен начин върху лист, поставен върху таблета като същевременно се получава и динамичната информация за всяка точка от траекторията на подписа, която се записва в създадената база данни за подписи.

Участниците в експериментите са 8 колеги от секция "ОСРО"- БАН с различна възраст и пол. Всеки от тях демонстрира индивидуален стил на подписване, така подписите на някои от участниците представляват неразбираеми щрихи, а на други четим и стегнат почерк.

Целта на протокола за събиране на подписи е описанието на процеса и детайлите по практическото събиране на образци на подписи, което се извършва с графичен таблет с помощта на разработеното за целта софтуерно приложение.

И така, броят на събраните собствени подписи е 80 (по 10 подписа от потребител), а този на умелите фалшификати - 72 (по 9 подписа за потребител).

#### 5.1.2 Бази данни за подписи SUsig

Базата данни за подписи SUsig [24] е предоставена от университета Sabanci University – Турция. Тя се състои от две части, съответно Visual и Blind. Първата съдържа подписи, събрани с помощта на таблет за подписи с вграден екран Interlink Elec. ePad. За всеки потребител се събрани 20 собствени и 10 подправени подписа, като истинските са събрани в две сесии. Броят на участниците е 84. За събирането на подписите от втората част (Blind) е използван графичният таблет Wacom Graphire2. С него са събрани по 8 или 10 собствени и 10 подправени подписа в една сесия за всеки участник. Броят на участниците е 89. За някои от потребителите броят на събраните собствени подписи е 8, а не 10 и поради това общият брой събрани подписи е 830. За всеки от подписите, в отделен текстов файл, се съдържа информация за координатите  $x$  и  $y$ , времеви отпечатък, ниво на натиск и индикатор за начало на шрих. Въз основа на предоставените характеристики на подписите, се пресмятат времетраенето и дължината (като брой точки) на всеки един от щрихите.

### 5.1.3 Диаграма на базата данни за подписи SUsig

## 5.2 Среда за разработка

## 5.3 Екрани

## Заклучение

В настоящата дисертационна работа е предложен комбиниран метод за разпознаване на он-лайн подписи, който е реализиран в софтуерно приложение. Разгледан е случая на верификация на подпис, поради все по-широкото му използване в обществената практика. Обхванат е целия процес по създаването на система за разпознаване по подписи с всички необходими за целта стъпки, включващи най-общо събиране на подписи, избор на признаци и създаване на модели за верификация. Изследвано е значението на типа подправени (умели и неумели) подписи както за избора на признаци, така и за самата верификация. Експериментите са проведени върху създадена за целта база от данни за подписи от осем участника, както и върху базата от данни за подписи SUsig, съдържаща подписи от 89 участника. Постигнатите резултати по време на експериментите демонстрират задоволителна точност на разпознаване от 98,46 %. Извършено е сравнение на получените резултати с класификатор НМ и класификатор по  $k$ -най-близки съседи. По такъв начин поставените в дисертационната работа цели задачи са изпълнени напълно.

Разработките и изследванията в областта на разпознаването на он-лайн подписи ще продължат и в бъдеще и ще бъдат насочени: 1) към усъвършенстване и подобряване на работата на системата като степен на автоматизация и точност; 2) тестването на системата върху различни бази данни и с използването на други признаци; 3) поради все по-широкото навлизане на такива системи в практиката и възможността за измами все по-често ще възниква въпросът за установяването на фалшификати. В тази връзка от голяма полза ще бъде добавянето на телевизионна камера, която ще отчита допълнително характеристиките и движението на ръката на подписващия се. Очертава се възможност и за съвместна изследователска работа с експерти от НИКК при МВР и участие в национални и международни проекти.

## Авторска справка

### Научно-приложните приноси са:

- Предложен е подход за минимизиране на признаково пространство чрез последователно прилагане на метода на корелационните плеяди и регресионния анализ. С прилагането му е получено минимално подмножество от признаци за всеки потребител, включен в база данни;
- Предложен е метод за създаване на класификатор като комбинация от индивидуални за всеки потребител класификатори;
- Предложен е обобщен мрежов модел на процеса за он-лайн верификация на подписи;
- Извършено е сравнение на точността на верификацията с използването на умели и неумели фалшификати в процеса на обучението;
- Извършено е сравнение на точността на верификацията по общи и индивидуални признаци.

### Приложни приноси:

- Разработена е база данни и протокол за събиране на он-лайн подписи;
- Разработен е прототип на софтуерна система, даваща възможност за: добавяне на потребител, търсене на даден потребител, събиране на подписи с графичен таблет, предварителна обработка на данните за подписи, извличане на признаци, избор на признаци, визуализация на собствени и фалшиви подписи, избор на модел на класификатор и обучението му;
- Създаден е речник на част от специфичните термини в разпознаването на образи на български и английски език.

### Благодарности

Изваквам своята сърдечна благодарност на научния си ръководител доц. д-р Георги Глухчев за компетентните напътствия, професионалните съвети и търпението, оказани при разработването на дисертационния труд.

Благодаря и на колегите от колектива на секция "Обработка на изображения и разпознаване на образи" с ръководител доц. д-р Димо Димов за предоставянето на образци от подписите им, необходими за провеждане на част от експериментите.

Благодаря на съпруга и близките ми за подкрепата и разбирането, които проявиха по време на моята работа.

Работата по дисертационния труд е осъществена с подкрепата на Проект: BG051PO001-3.3.04/40 по Оперативна програма "Развитие на човешки ресурси" на ЕСФ и МОМН - Република България.

### Литература

- [1] Plamondon, R., Lorette, G.: *Automatic signature verification and writer identification – the state of the art*. Pattern Recognition 22, pp.107–131, 1989.
- [2] J. Fierrez-Aguilar, L. Nanni, J. Lopez-Penalba, J. Ortega-Garcia, and D. Maltoni. *An online signature verification system based on fusion of local and global information*. In Proc. of IAPR Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA, pp. 523–532. Springer LNCS-3546, 2005.
- [3] A. Kholmatov and B. Yanikoglu. *Identity authentication using improved online signature verification method*. Pattern Recognition Letters, 26(15):2400–2408, 2005.
- [4] Chatterjee, S., Hadi, A.: *Regression Analysis by Example*. 4th Ed. New York, 2006.
- [5] D. Impedovo et al. *Automatic signature verification – The state of the art*, IEEE Transaction on Systems, Man and Cybernetics part C, pp. 609-635, Vol. 38, Issue 5, 2008
- [6] P. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [7] Jain A., Stan Li. *Encyclopedia of Biometrics*, Springer, 2009.
- [8] L. Likforman-Sulem, S. Garcia-Salicetti, J. Dittmann, J. Ortega-Garcia, N. Pavesic, G. Gluhchev, S. Ribaric, and B. Sankur, *Report on the hand and other modalities stated of the art*, Biometrics for Secure Authentication, 2005 .
- [9] LaMotte, L.R., Hocking, R.R.: *Computational Efficiency in the Selection of Regression Variables*. *Technometrics*. 12, pp. 83-93, 1970.
- [10] Boyadzhieva D., Gluhchev G., *Feature Set Selection for On-line Signatures using Selection of Regression Variables*, In: Proceedings of the 4th International Conference on Pattern Recognition and Machine Intelligence PReMI'11, pp. 440-445, 2011.
- [11] Neural Network Toolbox™ 7
- [12] G. P. Zhang, *Neural networks for classification: a survey*, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 30, pp. 451-462, 2000.
- [13] McCabe, Alan, Trevathan, Jarrod, and Read, Wayne. *Neural network-based handwritten signature verification*. Journal of Computers, 3 (8). pp. 9-22, 2008.

- [14] Hocking, R.R.: *The Analysis and Selection of Variables in Linear Regression*. Biometrics. 32, 1976.
- [15] Hocking, R.R., Leslie, R.n.: *Selection of the Best Subset in Regression Analysis.*, Technometrics. 9, pp. 531-540, 1967.
- [16] Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, *Regression analysis by example*, 4th ed, Wiley, 2007.
- [17] Mallows, C.L. *Some Comments on CP*. Technometrics 15 (4): pp. 661–675, 1973.
- [18] <http://www.microsoft.com/visualstudio/en-us/products/2010-editions/express>
- [19] <http://www.microsoft.com/sqlserver/en/us/editions/compact.aspx>
- [20] <http://www.microsoft.com/download/en/details.aspx?displaylang=en&id=20039>
- [21] [http://www.wacom-asia.com/intuos3/intuos3\\_630.html](http://www.wacom-asia.com/intuos3/intuos3_630.html)
- [22] <http://www.mathworks.com/products/matlab/>
- [23] <http://www.mathworks.com/matlabcentral/answers/22944-neural-network-design>
- [24] Alisher Kholmatov, Berrin A. Yanikoglu: *SUSig : an on-line signature database, associated protocols and benchmark results*. Pattern Anal. Appl. 12(3): 227-236, <http://biometrics.sabanciuniv.edu/SUSig>, 2009 .
- [25] LeCun, L. Bottou, G. Orr and K. Muller: *Efficient BackProp*, in Orr, G. and Muller K. (Eds), *Neural Networks: Tricks of the trade*, Springer, 1998.
- [26] Dougherty G., *Pattern Recognition and Classification*, Springer-Verlag New York Inc. ISBN: 9781461453222.
- [27] Kuncheva L.I. *Combining classifiers: Soft computing solutions*, in: S.K. Pal and A. Pal (Eds.) *Pattern Recognition: From Classical to Modern Approaches*, World Scientific Publishing Co., Singapore, pp.427-452, 2001.
- [28] Ink Data: <http://msdn.microsoft.com/en-us/library/ms811395.aspx>
- [29] Fawcett, T. *ROC graphs: Notes and practical considerations for researchers*, Tech Report HPL-2003-4, HP Laboratories. <http://www.hpl.hp.com/techreports/2003/HPL-2003-4.pdf>, 2003.
- [30] С.А. Айвазян, З.И. Бежаева, О.В. Староверов, Классификация многомерных наблюдений, Москва, Статистика, 240 стр., 1974.
- [31] Gluhchev G., M. Savov, O. Boumbarov, D. Vassileva. *A New Approach to Signature Based Authentication*, 2nd Int. Conf. on Biometrics, Seoul, pp. 594-603, 2007.
- [32] Savov M., G. Gluhchev. *Signature verification via Hand-Pen motion investigation*, Proc. Int. Conf. "Recent Advances in Soft Computing", Canterbury, 2006, pp. 490-495.
- [33] Nalwa, V.: *Automatic on-line signature verification*. Proc. IEEE 85, 215–239 , 1997.
- [34] J. Richiardi, H. Ketabdar, A. Drygajlo, *Local and Global Feature Selection for On- line Signature Verification*, IEEE 8th International Conference on Document Analysis and Recognition (ICDAR'05), vol.2, pp. 625 – 629, 2005.
- [35] Atanassov, K., G. Gluhchev, S. Hadjitodorov, A. Shannon, V. Vassilev, *Generalized Nets and Pattern Recognition*. KvB Visual Concepts Pty Ltd, Monograph No. 6, Sydney, 2003.
- [36] *A generalized net model of the process of handwriting identification*. In: [35].
- [37] Schmidt T., Rizzo V., Mery D.: *Dynamic Signature Recognition Based on Fisher Discriminant*. CIARP, volume 7042 of Lecture Notes in Computer Science, Springer, pp.433-442, 2011.
- [38] Abdulla Ali, *Offline Signature Verification using Radon Transform and SVM/kNN Classifiers*, ISSN 0136-5835, Вестник ТГТУ, Том 15. № 1, 2009.
- [39] Meenakshi K., Sargur S., Aihua X., *Offline Signature Verification And Identification Using Distance Statistics*, In: International Journal of Pattern Recognition and Artificial Intelligence, Vol. 18, No. 7, pp.1339-1360, 2004.
- [40] Gluhchev G., M. Savov, O. Boumbarov, D. Vassileva. *A New Approach to Signature Based Authentication*, 2nd Int. Conf. on Biometrics, pp. 594-603, 2007.
- [41] Savov M., G. Gluhchev. *Signature verification via Hand-Pen motion investigation*, Proc. Int. Conf. Recent Advances in Soft Computing, pp. 490-495, 2006.
- [42] Berrin A. Yanikoglu, Alisher Kholmatov: *Online Signature Verification Using Fourier Descriptors*. EURASIP J. Adv. Sig. Proc., 2009.

