FRI-ONLINE-1-CCT2-09

VOICE CLASSIFICATION BY ARTIFICIAL NEURAL NETWORKS WITH LM AND SCG ALGORITHMS [21]

Assoc. Prof. Ivelina Balabanova, PhD
Department of Communications Equipment and Technology,
Technical University of Gabrovo, Bulgaria
Phone: 0896 640 473
E-mail: ivstoeva@abv.bg

Chief Assist. Georgi Georgiev, PhD
Department of Communications Equipment and Technology,
Technical University of Gabrovo, Bulgaria
Phone: 0877 522 029
E-mail: givanow@abv.bg

Abstract: *In this paper the applicability of artificial neural networks with combination of Levenberg-Marquardt (LM) and Scaled Conjugate Gradient (SCG) training algorithms for identification of speakers by speech analysis have been studied. The voice classification process of participants is based on neural learning procedures by individual sound characteristics registered in speech activity. According to the training algorithms, the following criteria groups "Accuracy and Mean Squared Error (MSE)" for LM algorithm and "Accuracy and Cross-Entropy" in SCG algorithm were evaluated. The presented results consists of performance analysis, correct classifications/ misclassifications, ROC analysis, error levels between predicted and observed values.*

Keywords: *Speaker Identification, Sound Analysis, Artificial Intelligence, LM and SCG Algorithms, Accuracy.*

ВЪВЕДЕНИЕ

Моделирането, обработката и разпознаването на сигнали и реч се свързва с редица предизвикателства пред съвременните изследователи и разработчици на приложения и концептуални системи за гласов анализ, където се срещат различни по функционалност подходи на науката и инженерната практика (Iliev, M., Bedzhev, B., Bedzheva, M., & Kanchev, K., 2019; Iliev, M., Bedzhev, B., Bedzheva, M., & Yanakiev, P., 2020). Някои от тях най-често се базират на преобразуването на речта в текст, а процесът на разпознаване се свежда до извличане на информация от отделни морфеми, думи и изрази. Основни инструменти, използвани за гласово моделиране, обработка и анализ, са изчислителните ресурси на изкуствените невронни мрежи. Тук главно могат да бъдат посочени архитектури на базата на радиално базисни функции или придобилите през последните години популярност конволюционни изкуствени невронни мрежи (Gevaelt, W., Tsenov, G., & Mladenov, J., 2019; Kubanek, M., Bobulski, J., & Kulowik, J., 2019). Процедурите се състоят в предварителна сигнална обработка, извличане на информативни признаци, формиране на база данни, класификация и оценка на производителността. Относно целите на гласовата идентификация се използват специализирани автоматизирани системи на основата на безжични комуникации, web интегрирани решения и други (Sigh, D., & Thakur, A., 2013; Rankovska, V., & Rankovski, S., 2017; Rankovska, V., 2019; Washani, N., & Sharma, S., 2015).

В доклада се разглежда задачата за гласова идентификация на реални лица (група от шест лектора, от които $4^{-ри}$ жени и $2^{-ма}$ мъже) чрез звуков анализ на речта и изкуствен интелект на основата на LM и SCG обучение. Изследването се базира на снемане на група от показатели, между които LZE, LZeq, LZF, LZS, LZI, LAE, LAeq, LAF, LAS, LAI, LCE, LCeq, LCF,LCS и LCI при анализ на реч във фиксиран времеви интервал (60 секунди – 600 ms) чрез

[21] Докладът е представен на on-line сесия на 13 ноември 2020 с оригинално заглавие на български език: ГЛАСОВА ИДЕНТИФИКАЦИЯ ПОСРЕДСТВОМ ИЗКУСТВЕНИ НЕВРОННИ МРЕЖИ С LM И SCG ОБУЧЕНИЕ

софтуерен звуков анализатор и синтез на невронни мрежи при различни функции на активация в изходните слоеве.

ИЗЛОЖЕНИЕ

Синтез на изкуствени невронни мрежи на основата на LM обучение

Реализирано е обучение на изкуствени невронни мрежи при зададени линейна, тангенс-сигмоидална и логаритмична-сигмоидална активационна функция на изхода по LM алгоритъм за гласова идентификация. Резултатите от проведените дейности при оценка на качеството на моделите според постинатите критерии „Точност“ и „Средноквадратична грешка“ при промяна на скритите неврони в интервала от 5 до 15 са показани от таблица 1 до таблица 3.

Таблица 1. Резултати при синтез на изкуствени невронни мрежи с LM обучение при линейна изходна активация

№	Скрити неврони	Точност, %	Средноквадратична грешка
1	5	95.4	0.0142
2	6	96.1	0.0125
3	7	99.4	0.0098
4	8	97.4	0.0104
5	9	98.9	0.0098
6	10	99.1	0.0111
7	11	99.3	0.0098
8	*12*	*99.8*	*0.0087*
9	13	99.1	0.0093
10	14	99.3	0.0095
11	15	98.1	0.0117

Таблица 2. Резултати при синтез на изкуствени невронни мрежи с LM обучение при тангенс-сигмоидална изходна активация

№	Скрити неврони	Точност, %	Средноквадратична грешка
1	5	97.8	0.0102
2	6	97.4	0.0100
3	7	98.7	0.0089
4	8	94.4	0.0106
5	9	98.7	0.0082
6	10	97.2	0.0100
7	11	98.1	0.0102
8	12	97.4	0.0107
9	13	99.3	0.0066
10	*14*	*99.8*	*0.0069*
11	15	97.8	0.0092

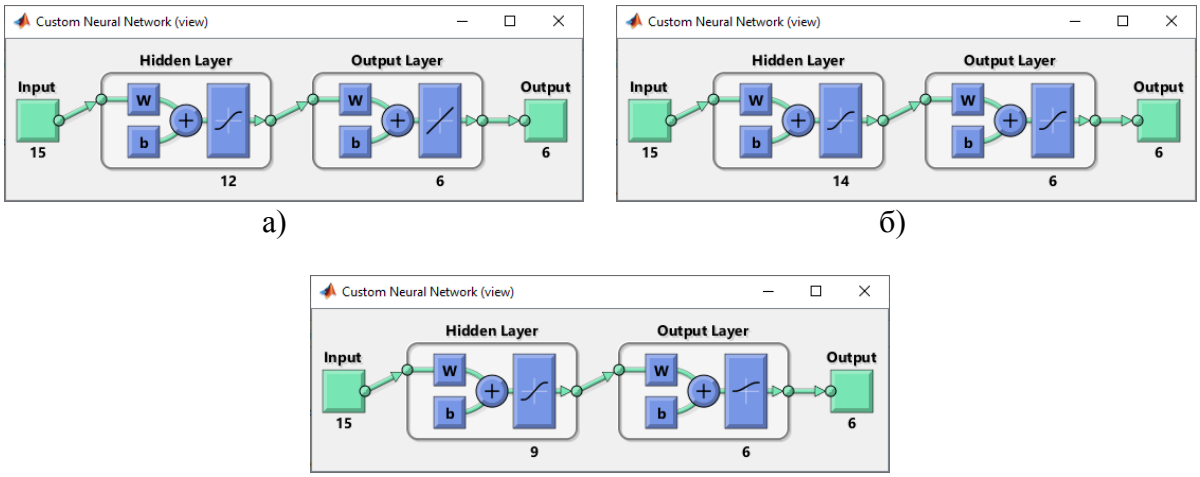
Относно използването на линеен тип на активация в изходните мрежови слоеве са установени нива на точност над 95.0 %. Вариациите на показателя попадат между 95.4 % и 99.8 %, съответно при 5 и 12 междинни неврона. Максималната MSE от порядъка на 0.0142, както и регистрираната най-ниска стойност на грешката 0.0087, са получени също за посочените невронни единици в междинния слой. Подобна тенденция се наблюдава по отношение на посочените показания за индикаторите при тангенс-сигмоидална функция като тук най-ниската точност 94.4 % е констатирана при 8 скрити неврона. Достигната е еднаква

степен на максимална точност (99.8 %) при архитектура с 14 неврона в междинния слой, за която е наблюдавана по-ниска средноквадратична грешка MSE = 0.0069 - по-добра спрямо линейния тип на активация. Същият извод може да бъде направен и във връзка с максималната средноквадратична грешка 0.0107, получена при архитектура с 12 скрити неврона.

Значително влошаване на индикаторите от оценка на качеството се откроява при логаритмична-сигмоидална функция, особено изразено по отношение на MSE. За анализирания случай грешката се изменя в диапазона от 0.2329 при 9 до 0.2500 при 5 неврона в междинния слой. Тук регистрираната минимална точност се равнява едва на 16.7 % (при 5 скрити неврона), а установената най-висока съответно 47.8 % (при 9 скрити неврона). Това определя последният приложен тип на изходна активация като неприложим за целите на анализа.

Таблица 3. Резултати при синтез на изкуствени невронни мрежи с LM обучение при логаритмична-сигмоидална изходна активация

№	Скрити неврони	Точност, %	Средноквадратична грешка
1	5	16.7	0.2500
2	6	34.4	0.2357
3	7	37.2	0.2359
4	8	46.7	0.2338
5	*9*	*47.8*	*0.2329*
6	10	40.4	0.2387
7	11	37.6	0.2426
8	12	35.0	0.2402
9	13	38.1	0.2428
10	14	35.2	0.2433
11	15	25.7	0.2432



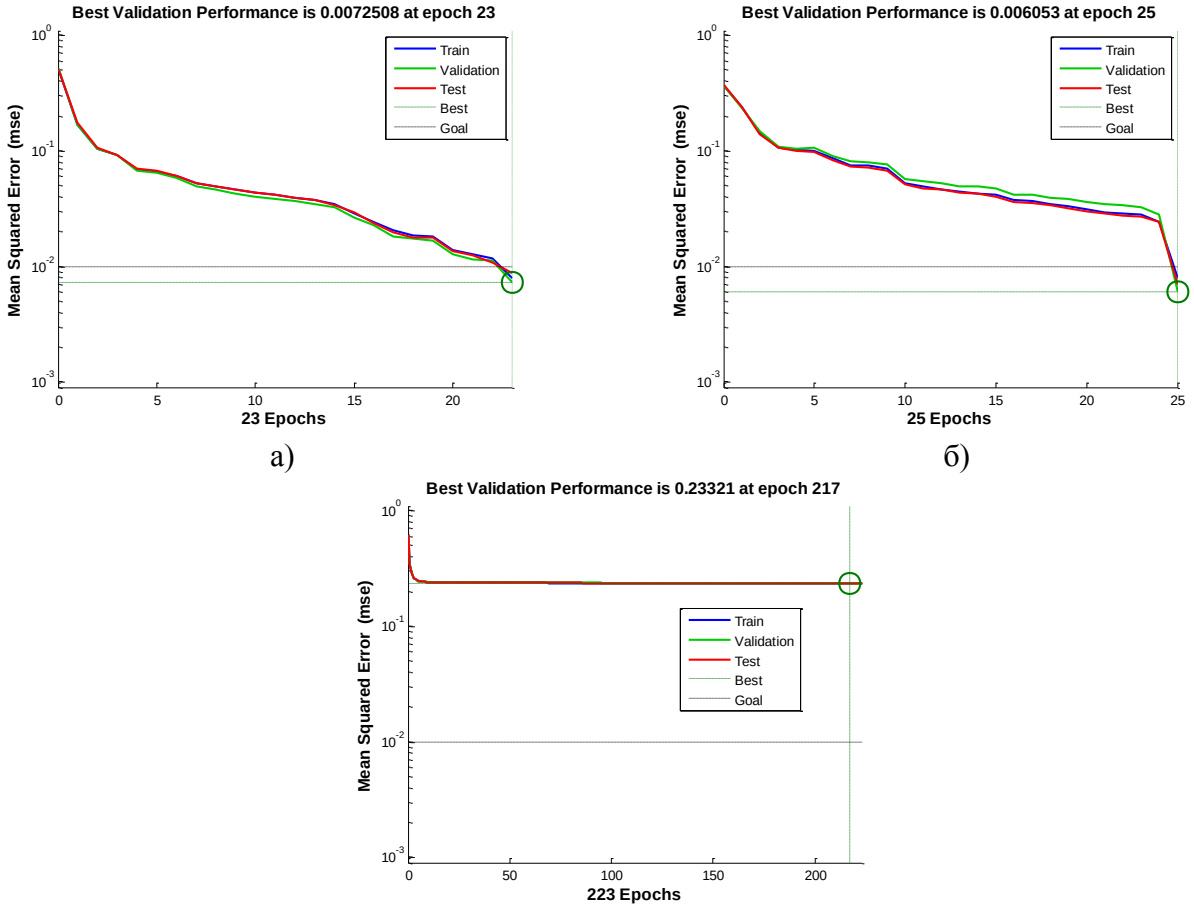
а) б)

в)

Фиг. 1. Архитектури на невронни мрежи за гласова идентификация на лица с LM обучение при а) линейна, б) тангенс-сигмоидална и в) логаритмична-сигмоидална изходна функция

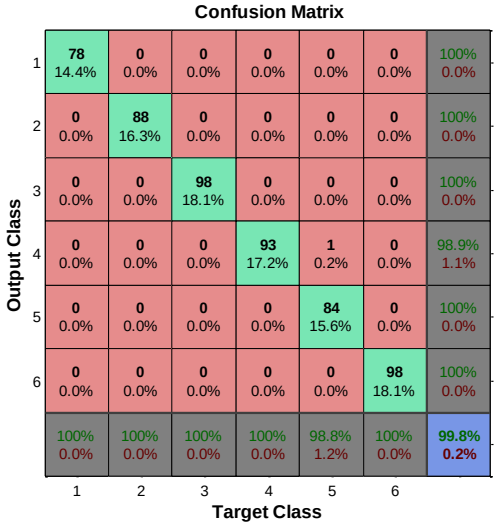
Синтезираните крайни архитектури на изкуствени невронни мрежи за идентификация на физически лица съобразно звуков анализ при произнасяне на реч са представени на фиг. 1. При използване на посочения обучаващ алгоритъм с най-висока степен на адекватност е определен тангенс-сигмоидалният тип на активация на изходния слой.

Относно синтезираните архитектури са дадени осцилограми на средноквадратичните грешки за процесите на мрежово обучение, валидиране и тестване, за които се забелязва обща тенденция по отношение на характера на изменение (фиг. 2). Не се наблюдават индикации за възникнали проблеми в хода на процесите. Визуално правят впечатление разликите по отношение на високата степен на разграничаване между нивата на целевата и достигната най-добра MSE, както и значително по-големия брой обучаващи итерации, констатирани при мрежата с изходна логаритмичната-сигмоидална функция. Постигнати са най-добри производителности от верификация, както следва 0.0072508 при 23$^{-та}$, 0.006053 при 25$^{-та}$ и 0.23321 в обхвата на 217 цикъла за обучение.

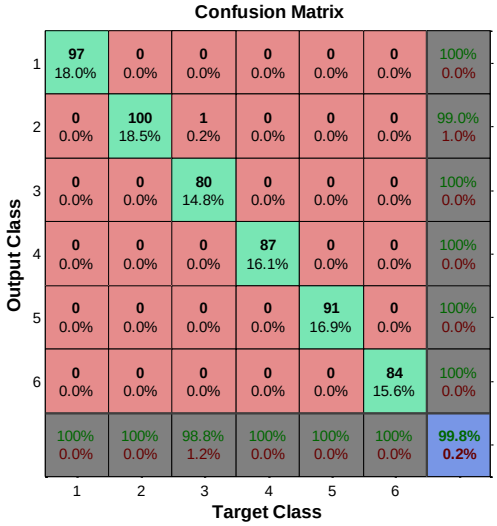


Фиг. 2. MSE при селектирани архитектури на невронни мрежи за гласова идентификация на лица с LM обучение при а) линейна, б) тангенс-сигмоидална и в) логаритмична-сигмоидална изходна функция

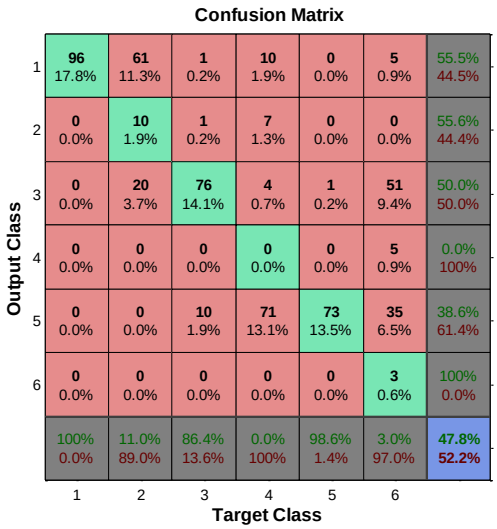
Фигура 3 илюстрира разпределението на данните от тестовите поднабори във връзка с реализираните коректни и некоректни класификации. Относно мрежите с линейна и тангенс-сигмоидална активация на изхода са установени едва по един еталон с неправилно определена принадлежност, съответно от пета и трета изходна група. По-различен начин е разглежданото разпределение при активационния тип с най-ниска степен на адекватност, където при речева диагностика е констатирана най-висока успеваемост в нисходящ ред при първо, пето, трето, второ, шесто и четвърто лице. Относно последният лектор е налице пълна некоректност при разпознаване на тестови речеви фрагменти.



а)

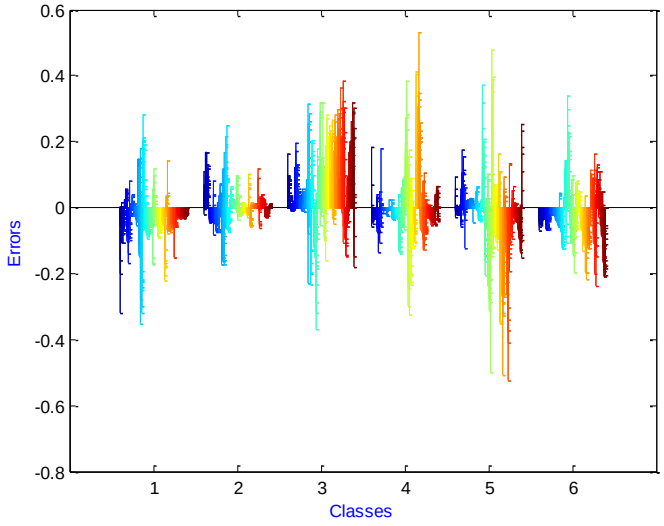


б)

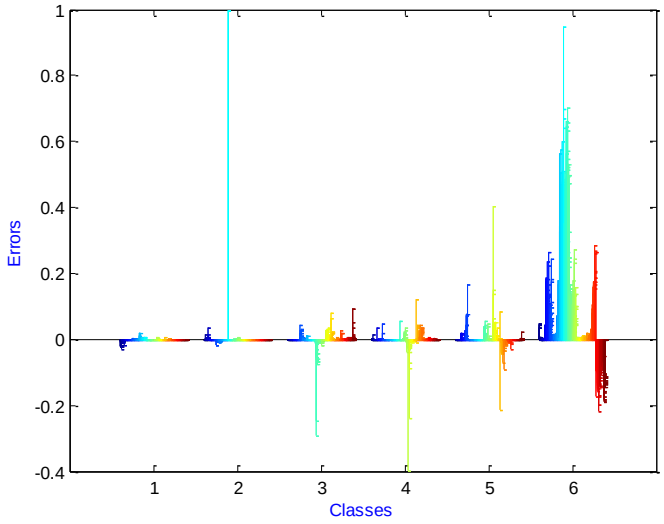


в)

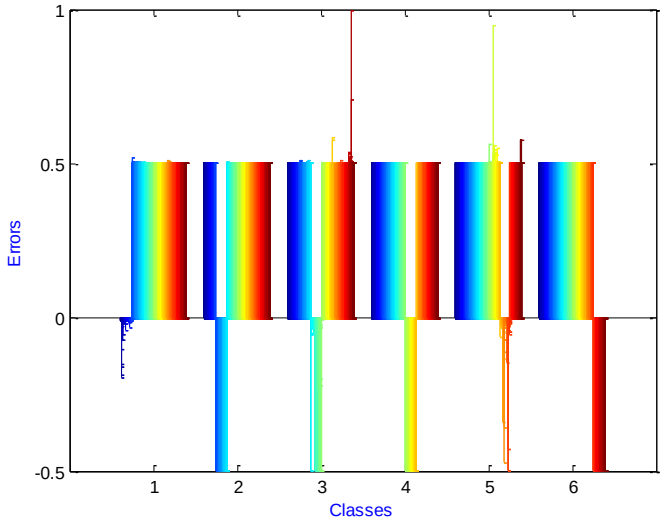
Фиг. 3. Класификационни матрици относно невронни мрежи за гласов анализ на лица с LM обучение при а) purelin, б) tansig и в) logsig изходни типове на активация



а)



б)



в)

Фиг. 4. Абсолютни мрежови грешки при синтезирани архитектури на невронни мрежи за гласова идентификация на лица с LM обучение при а) линейна, б) тангенс-сигмоидална и в) логаритмична-сигмоидална изходна функция

Вариациите на абсолютните мрежови грешки (разликите между калкулираните и целевите стойности за изследваните невронни архитектури) са онагледени на фиг. 4. При линейна изходна активационна функция са констатирани минимален -0.5239 и максимален 0.5307 праг на изменение. При модела с тангенс-сигмоидална активация на изхода указаните прагове се равняват на -0.3987 и 0.9998. Макар и наличието на ясно изразени и ограничени по количество пикове при втора и шеста класификационна група, както беше посочено в по-горната част от изследването, тук е установена най-ниска средноквадратична грешка 0.0069 за цялостния вариационен диапазон. За мрежата с логаритмична-сигмоидална функция ясно се виждат в значителен процент преобладаващите грешки за данните от тестовата извадка при ниво 0.5, които са в пъти по-високи спрямо получените при почти всички лица - обект на гласова анализ с помощта на невронен модел при tansig дефиниран тип на изхода.

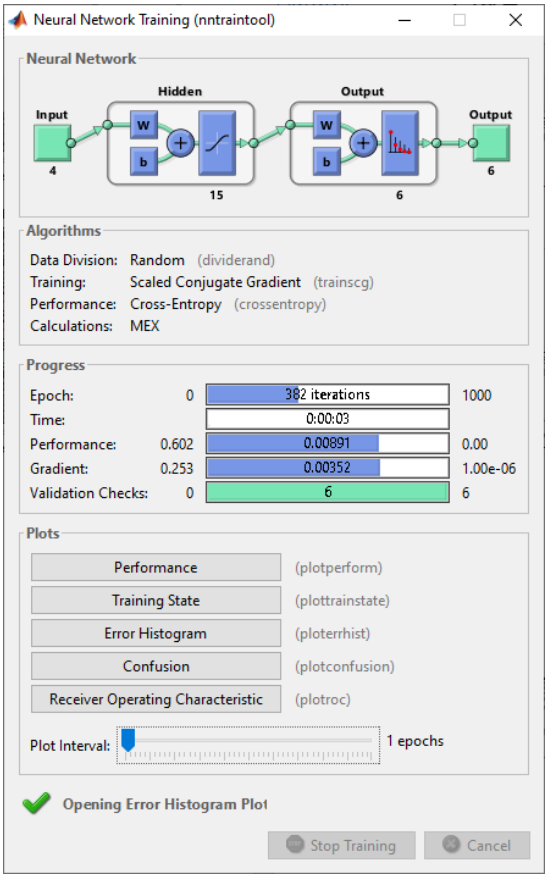
Подбор на изкуствена невронна мрежа на базата на SCG обучаващ алгоритъм

След изследване на приложимостта на Levenberg-Marquardt се премина към манипулации с използване на Scaled Conjugate Gradient алгоритъм на обучение, където бяха въведени оценка на специфичен критерий Cross-Entropy от тестване и softmax активационен тип в изходния слой на мрежите. Точността беше запазена като базисен анализиран показател при синтез на невронен модел за гласово разпознаване и класификация на реални физически лица. Първоначално при обучение бяха приложени идентични звукови показатели при анализ на речта, но се установиха сравнителни и незадоволителни нива на точност, което наложи изършването на допълнителни процедури по паралелно редуциране и подбор на входни променливи. В резултат от направените дейности бяха селектирани следните четири информативни признака – LZE, LZeq, LZF и LZS.

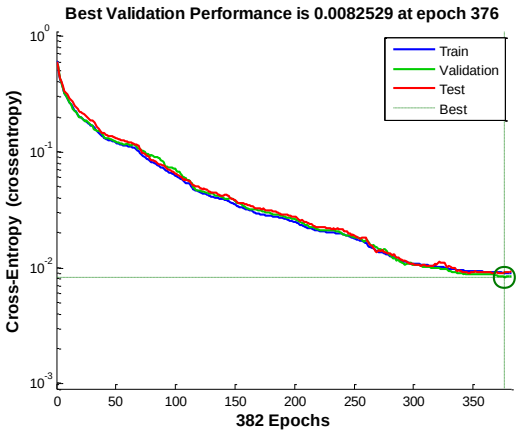
В таблица 4 са обобщени резултатите от оценка на невронни архитектури при промяна на скритите неврони в границите от 5 до 20. Най-ниска точност 85.4 % е намерена при 6 междинни неврони, докато максимално показание на индикатора 99.5 % беше установено при 15 изчислителни единици в скрития слой. Минимална Cross-Entropy = 7.13753e-0 е получена при долна граница на междинните неврони. Осносо селектирания молел с най-добра ефективност, чиито ход от обучение може да бъде наблюдаван на фиг. 5, е достигната стойност на индикатора 16.95265e-0.

Таблица 4. Резултати при синтез на изкуствени невронни с SCG обучение

№	Скрити неврони	Точност, %	Cross-Entropy
1	5	89.5	10.46477e-0
2	6	85.4	7.13753e-0
3	7	93.8	10.49458e-0
4	8	88.8	9.27192e-0
5	9	97.9	14.48791e-0
6	10	96.4	14.01021e-0
7	11	98.0	15.18417e-0
8	12	96.9	13.27742e-0
9	13	97.2	13.84981e-0
10	14	95.1	11.39161e-0
11	*15*	*99.5*	*16.95265e-0*
12	16	95.9	14.08092e-0
13	17	99.1	14.56172e-0
14	18	98.5	15.68848e-0
15	19	97.2	14.27011e-0
16	20	97.7	14.10622e-0



Фиг. 5. Ход от обучение на селектираната невронна мрежа при SCG алгоритъм

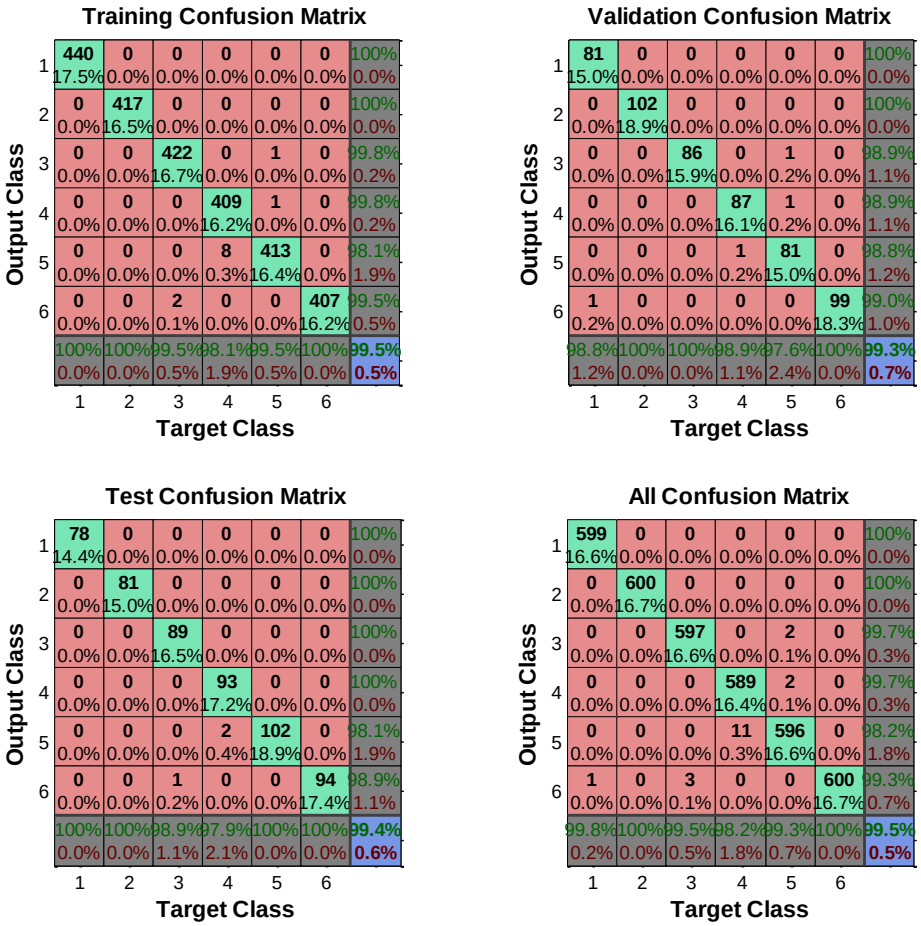


Фиг.6. Cross-Entropy при синтезираната невронна мрежа по SCG алгоритъм

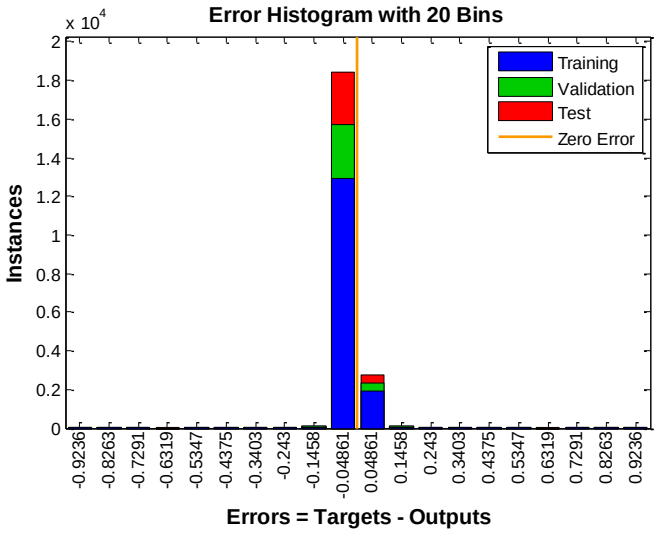
Фигура 6 визуализира промяната на индикатора Cross-Entropy във връзка с обучаващите, валидиращите и тестовите процедури. Най-висока производителност 0.0082529 е достигната при 376-та итерация от обучение, което се преустановява при 382-ри цикъл. Указаните характеристики притежават почти идентичен и много близък характер на изменение при проследяване относно обучаващите итерации.

На фиг. 7 са дадени матриците на коректни и некоректни класификации поотделно за основните мрежови процеси и крайната обобщаваща матрица като може да бъде направена следната интерпретация на резултатите. При обучение с правилно определена принадлежност са всички звуковите еталони относно първи, втори и шести лектор. По отношение на процедурите от верификация това може да се каже съответно при втора, трета и шеста изходна група. Съобразно тестовите процедури пълна коректност се дефинира относно първи, втори, пети и шести лектор. Общата класификационна матрица показва пълна коректност при втора и шеста изходна група.

Разположението на грешките в непосредствена близост до нулевата линия на хистограмата фиг. 8 за голям дял от данните от входния информационен набор потвърждава пригодността на синтезираната невронна архитектура за целите на изследването. Особеност тук е, че сами по себе си грешките нямат числова, а вероятностна интерпретация. Този факт се налага от спецификата на използваната softmax функция в изходния слой на анализираните невронни модели. За конкретния случай е констатиран минимален ортицател праг на грешките относно данните, участващи при обучение, валидине и тестване, както следва -0.048061.



Фиг. 7. Класификационни матрици за избраната с най-добри параметри невронна мрежа при SCG алгоритъм



Фиг. 8. Хистограма на грешките относно синтезираната мрежа при SCG алгоритъм

ЗАКЛЮЧЕНИЕ

Резултатите от направеното изследване дават основание предложения подход за обработка и анализ на реч на основата на извлечени звукови характеристики и последващо обучение на изкуствени невронни мрежи с цел идентификация на реални физически лица успешно да бъде внедрен в различни типове системи. Такива могат да бъдат системи за

гласов анализ и разпознаване, контрол на достъпа с образователна, търговска, корпоративна, административна, социална или индустриална насоченост. Във връзка с модификация на подхода се предвижда прилагане на преобразование на Фурие при различен тип на прозоречната функция върху регистрирани гласови активности и изследване на ефекта по отношение на успеваемостта на гласово разпознаване.

REFERENCES

Iliev, M., Bedzhev, B., Bedzheva, M., & Kanchev, K. (2019). *An algorithm for synthesis of mismatched filters for processing of aperiodic phase manipulated signals*. Proceedings of 16th Conference on Electrical Machines, Drives and Power Systems (ELMA 2019), 6 – 8 June 2019, Varna, Bulgaria.

Iliev, M., Bedzhev, B., Bedzheva, M., & Yanakiev, P. (2020). *A method for synthesis of nearly ideal phase manipulated signals*. Proceedings of the 2020 IEEE International Conference on Information Technologies (InfoTech-2020), 17-18 September 2020, St. St. Constantine and Elena, Bulgaria.

Gevaelt, W., Tsenov, G., & Mladenov, J. (2019). Neural networks used for speech recognition. *Journal of Automatic Control*, 20, 1-7.

Kubanek, M., Bobulski, J., & Kulowik, J. (2019). A method of speech coding for speech recognition using a convolutional neural network. *Summetry*, 11(1185), 1-12.

Sigh, D., & Thakur, A. (2013). Voice recognition wireless home automation system based on zigbee. *Journal of Electronics and Communication Engineering*, 22(1), 65-75.

Rankovska, V., & Rankovski, S. (2017). *A short survey on wireless interfaces in embedded systems*. Proc. of International Scientific Conference on Information, Communication Energy Systems and Technologies, 28-30 June 2017, Nis, Serbia.

Rankovska, V. (2019). *Web-based monitoring and control in embedded systems teaching*. Proc. of XXVII International Scientific Conference Electronics, 12-14 September 2019, Sozopol, Bulgaria.

Washani, N., & Sharma, S. (2015). Speech recognition system: A review. *International Journal of Computer Applications*, 115(18), 7-10.