

استفاده از شبکه های تجمیع خبرگان به منظور طبقه بندی سیگنال مغزی افراد سالم و

بیماران صرعی

سنا آموزگار^۱، محمد پویان^۲ و الیاس ابراهیم زاده^۳

۱- کارشناس ارشد، دانشکده مهندسی پزشکی - دانشگاه شاهد-تهران-ایران-

Sana_Amozegar@yahoo.com

۲- دانشیار، دانشکده مهندسی پزشکی - دانشگاه شاهد-تهران-ایران

Pooyan@shahed.ac.ir

۳- کارشناس ارشد، دانشکده مهندسی پزشکی - دانشگاه شاهد-تهران-ایران

Elias.Ebrahimzadeh@yahoo.com

چکیده: EEG یکی از مهمترین و رایجترین مراجع برای مطالعه عملکرد مغز و اختلالات نورولوژیک است. به همین دلیل، تشخیص تغییرات EEG توسط سیستم‌های خودکار، موضوعی است که برای سال‌های متوالی تحت مطالعه است. از آنجا که در هر سیستم دسته بندی، صحت تصمیم گیری از اهمیت ویژه ای برخوردار است، لذا نیازمند وجود روش های طبقه بندی بهتر برای سیگنال مغزی هستیم. در این تحقیق، به دنبال ارائه یک سیستم هوشمند مرکب برای بهبود صحت طبقه بندی سیگنال مغزی (EEG) به گروه های مختلف هستیم. این مقاله استفاده از روش تجمیع خبرگان (Mixture of Experts (ME)) برای بهبود تفکیک سیگنال های EEG افراد سالم و بیماران صرعی را نشان داده و صحت طبقه بندی آن را ارزیابی کرده است. تصمیم گیری در دو مرحله انجام شد: (۱) استخراج ویژگی به وسیله روش های مختلف بردار ویژه و (۲) طبقه بندی با استفاده از طبقه بندی کننده های آموزش دیده شده توسط ویژگی های استخراج شده. ورودی های این سیستم هوشمند از ویژگی های مرکب، که متناسب با ساختار شبکه انتخاب شده، تشکیل شده اند. در این مطالعه سه روش مبتنی بر بردارهای ویژه (PISARENKO MUSIC, MINIMUM NORM) برای تولید تخمین طیف چگالی توان (PSD) انتخاب شدند. پس از پیاده سازی ME و آموزش آن روی ویژگی های مرکب، نشان داده که این روش می تواند به صحت طبقه بندی بالایی برسد. از این جهت، جداسازی سیگنال های مغزی بیماران صرعی در حالات مختلف و افراد سالم با صحت بالایی امکان پذیر است. از طرفی، از آنجا که مناسب بودن عملکرد شبکه عصبی به اندازه دسته های آموزش و تست بستگی دارد، در این مطالعه با تقسیم مشاهدات به سه گروه آموزش (۷۰٪)، آزمون (۲۰٪) و اعتبار سنجی (۱۰٪) موجبات بهتر آموزش دیدن شبکه فراهم شده و در نهایت صحت طبقه بندی تا ۹۹٫۵۰٪ افزایش یافته است.

واژه های کلیدی: سیگنال EEG، صرع، طبقه بندی، تجمیع خبرگان، شبکه های خبره.

۱- مقدمه

عملکرد مغز و اختلالات نورولوژیک است. به همین دلیل، سیستم‌های خودکار برای تشخیص تغییرات EEG برای سال‌های متوالی تحت مطالعه هستند. پردازش اتوماتیک می تواند به سه بخش پردازشگر مستقل تقسیم شود: پیش پردازش، استخراج یا انتخاب ویژگی و طبقه بندی. استفاده از سیگنال و یا تصویر، میانگین گیری، حذف نویز، گذاشتن حدآستانه، پیدا کردن لبه و بهبود کیفیت سیگنال یا تصویر،

EEG یکی از مهمترین و رایجترین مراجع برای مطالعه

۱ تاریخ ارسال مقاله : ۱۳۹۱/۰۸/۴۰

تاریخ پذیرش مقاله : ۱۳۹۲/۰۱/۱۸

نام نویسنده مسئول : الیاس ابراهیم زاده

نشانی نویسنده مسئول : تهران - اتوبان خلیج فارس -

دانشگاه شاهد - دانشکده مهندسی پزشکی

استفاده از شبکه های تجمیع خبرگان به منظور طبقه بندی سیگنال مغزی افراد سالم و بیماران صرعی

روش اعتبار سنجی^۲ برای انجام اعتبارسنجی، داده ها به ۷۰-۲۰ تقسیم بندی شده اند، به این ترتیب که ۷۰٪ آنها برای آموزش، ۲۰٪ برای تست و ۱۰٪ برای اعتبارسنجی استفاده شده اند. این امر باعث می شود شبکه آموزش بهتری ببیند و همچنین، شبکه از over fit شدن مصون بماند. در واقع، از این طریق صحت آزمون افزایش می یابد که این امر در حالت عدم حفظ کردن دادگان رخ داده، و توانمندی بهتر آموزش در شبکه را نشان می دهد.

۲- انتخاب داده

الگوریتم ارائه شده بر روی مجموعه داده های موجود در پایگاه http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3&changelang=3 ارزیابی شده است.

به غیر از الکترودهای مختلف ثبت برای EEG بیرونی و درونی، همه پارامترهای دیگر ثبت ثابت بودند.

پنج دسته (A تا E) هر کدام شامل ۱۰۰ سیگنال تک کاناله با طول ۲۶،۳ ثانیه در این مطالعه استفاده شد. این سیگنالها از ثبت های دنباله دار EEG انتخاب و بریده شدند. دسته A و B ترکیبی از سیگنال گرفته شده از ثبت های EEG سطحی است که بر روی پنج نفر داوطلب سالم با استفاده از جاگذاری استاندارد الکتروود (طرح محل الکترودهای سطحی مربوط به سیستم جهانی ۱۰-۲۰) انجام شده بود. داوطلبان در حالت آرام و بیدار با چشمان باز (A) و چشمان بسته (B) بودند. دسته C و D و E از آرشیو تشخیص قبل از عمل به دست آمده. EEG پنج بیمار انتخاب شده که ناحیه صرعی^۳ همه آنها مشخص بود. سیگنال دسته D از داخل ناحیه صرعی^۳ ثبت شده است و دسته C از ناحیه هیپوکامپ^۴ از نیمکره مقابل ثبت شده، دسته C و D از وقفه های حمله گرفته شده، در حالی که دسته E فقط شامل فعالیت حمله صرع است.

همه سیگنال های EEG با سیستم ۱۲۸ کاناله و با ولتاژ میانگین مشترک ثبت شدند. پس از تبدیل ۱۲ بیتی آنالوگ به دیجیتال، داده در دیسک ذخیره سازی داده به صورت

عملیات عمده در قسمت پیش پردازش هستند. ویژگی ها به وسیله بخش استخراج ویژگی به دست می آیند [۱]. بخش انتخاب ویژگی یک مرحله اختیاری است، که به وسیله آن فقط اندازه بردار ویژگی کوچک می شود و بخش طبقه بندی کننده آخرین مرحله در تشخیص اتوماتیک است [۲ و ۳].

روش های استخراج ویژگی به دو زیر گروه تقسیم می شود (۱): مشخصه های آماری (۲) توصیفات بیان کننده داده ها.

روش های مبتنی بر بردارهای ویژه، به عنوان یک روش استخراج ویژگی، برای تخمین فرکانس و توان سیگنال از اندازه گیری های آغشته به نویز استفاده می شوند. این روش ها بر پایه تجزیه ماتریس کوریلیشن سیگنال آغشته به نویز هستند. حتی وقتی نسبت سیگنال به نویز (SNR) پایین باشد، روش های مبتنی بر بردارهای ویژه یک طیف فرکانسی با رزولوشن بالا تولید می کنند. در این مطالعه، سه روش مبتنی بر بردارهای ویژه (PISARENKO, MUSIC, MINIMUM NORM)، برای تولید تخمین طیف چگالی توان^۱ (PSD) انتخاب شد. مدل پیشنهادی در این مقاله از سه ماژول اصلی تشکیل شده است: (۱) استخراج کننده ویژگی که بردار ویژگی را از سیگنال EEG تولید می کند؛ (۲) انتخاب کننده ویژگی و (۳) طبقه بندی کننده که داده های ورودی را به کلاس های مجزا طبقه بندی می کند [۴ و ۵].

سیگنال EEG تحت مطالعه شامل پنج دسته است: (۱) سیگنال سطحی ثبت شده از داوطلبان سالم با چشم های باز؛ (۲) سیگنال سطحی ثبت شده از داوطلبان سالم با چشم های بسته؛ (۳) سیگنال ثبت شده عمقی از بیماران صرعی در طول وقفه های بین حملات صرع از درون ناحیه ای که حمله را ایجاد می کند؛ (۴) سیگنال ثبت شده عمقی از بیماران صرعی در طول وقفه های بین حملات صرع از بیرون ناحیه ای که حمله را ایجاد می کند و (۵) ثبت عمقی حملات صرعی است.

در این مقاله، شبکه پیشنهادی ME برای آموزش و جداسازی دادگان انتخاب شده است. از طرفی، با توجه به

پس با استفاده از این فرمول ماتریس اتوکورلیشن (R) تولید می‌شود.

اگر a یک بردار ویژه ماتریس R باشد، معادله ۲ می‌تواند به صورت زیر بیان شود:

$$\hat{R}a = SPS^*a + \sigma^2 a \quad (4)$$

که a به صورت $[a_0; a_1; \dots; a_m]$ است. روش پیسارنکو فقط بردار ویژه مربوط به مقدار ویژه مینیمم را برای ساختن چند جمله‌ای (۱) استفاده می‌کند و طیف را محاسبه می‌نماید. بنابراین، روش پیسارنکو، a را طوری می‌یابد که $S^*a = 0$ شود. پس بردار ویژه a می‌تواند این گونه فرض شود که به زیر فضای نویز کشیده شده و معادله (۴) به حالت زیر خلاصه می‌شود:

$$\hat{R}a = \sigma^2 a \quad (5)$$

با این محدودیت که $a^*a = 1$ در جایی که δv^2 توان نویز است، که در روش پیسارنکو همان مقدار ویژه مینیمم مربوط به بردار ویژه a است. در اصل، با فرض سفید بودن نویز، مقدار ویژه همه زیر فضاهای نویزی باید با هم برابر باشند [۱]:

$$\lambda_1 = \lambda_2 = \dots = \lambda_k = \sigma^2 \quad (6)$$

که λ_i مقادیر ویژه زیر فضاهای نویزی را بیان می‌کند و $i=1, 2, \dots, k$ بیان کننده بعد زیر فضای نویزی است [۱]. روش پیسارنکو^۵ از بردار مربوط به مقدار ویژه مینیمم، PSD را از خروجی مطلوب چند جمله‌ای تعیین می‌کند [۱]:

$$P_{\text{Pisarenko}}(f) = \frac{1}{|A(f)|^2} \quad (7)$$

۳-۲- روش MUSIC

روش MUSIC یکی از تخمین‌گرهای فرکانسی زیر فضاهای نویزی است که توسط Schmidt (1986) ارائه شد و اثر صفرهای کاذب^۶ را با میانگین‌گیری روی همه طیف‌های بردارهای ویژه مربوط به زیر فضاهای نویزی از بین می‌برد. حاصل PSD به دست آمده توسط این روش به

پیوسته با نرخ نمونه برداری ۱۷۳،۶۱ هرتز نوشته شدند.

۳- روش‌های بردار ویژه برای آنالیز طیفی

۳-۱- روش پیسارنکو^۵

روش پیسارنکو^۵ که در سال ۱۹۷۳ بیان شد، مخصوصاً برای تخمین PSD هایی که شامل پیک های تیز در فرکانس‌های مورد انتظار هستند، مفید است. چند جمله‌ای $A(f)$ که صفرهای آن در دایره واحد است، می‌تواند برای تخمین PSD استفاده شود [۱]:

$$A(f) = \sum_{k=0}^m a_k e^{-j2\pi f k} \quad (1)$$

که در آن $A(f)$ خروجی مطلوب چند جمله‌ای است، a_k ضرایب چند جمله‌ای و m ترتیب فیلترهای ویژه را نشان می‌دهد. چند جمله‌ای به صورت ماتریس اتوکورلیشن سیگنال ورودی نیز می‌تواند نشان داده شود. فرض کنید که نویز، سفید باشد [۱]:

$$R = E\{x(n)^* \cdot x(n)^T\} = SPS^* + \sigma^2 I \quad (2)$$

که $x(n)$ سیگنال مورد نظر، S بیان کننده ماتریس جهت سیگنال با ابعاد $(m+1) \times L$ و بعد زیر فضای سیگنال است، R ماتریس اتوکورلیشن با ابعاد $(m+1) \times (m+1)$ است، p ماتریس توان سیگنال با ابعاد $L \times L$ است، σ^2 بیان کننده توان نویز، $*$ بیان کننده complex conjugate، I ماتریس یک، $\#$ بیان کننده complex conjugate، T و transpose بیان کننده ترانپوز است. S ، ماتریس جهت سیگنال، به صورت زیر بیان می‌شود: $S = [Sw_1 Sw_2 Sw_3 \dots Sw_L]$ که w_1, w_2, \dots, w_L بیان کننده فرکانس‌های سیگنال اند:

$$Sw_i = [1 \ e^{jw_i} e^{2jw_i} \dots \ e^{jm w_i}] \ , \ i = 1, 2, \dots, L$$

در طرح های عملی متداول است که تخمین ماتریس اتوکورلیشن را از دستور زیر محاسبه کنند:

$$\hat{R}(k) = \frac{1}{N} \sum_0^{N-1-k} x(n+k) \cdot x(n) \quad k = 0, 1, \dots, m$$

صورت زیر بیان می شود [۱]:

$$P_{\text{MUSIC}}(f) = \frac{1}{\frac{1}{K} \sum_{i=0}^{K-1} |A_i(f)|^2} \quad (8)$$

که بعد از k بعد از فضای نویزی را تعیین می کند، و $A_i(f)$ خروجی مطلوب چند جمله ای مربوط به همه بردارهای ویژه را تعیین می کند [۱].

۳-۳- روش مینیمم نرم

علاوه بر روش های MUSIC و Pisarenko روش مینیمم نرم هم برای جداسازی صفرهای کاذب از صفرهای حقیقی پیاده شد. PSD از روش مینیمم نرم می تواند این گونه تخمین زده شود [۱]:

$$p_{\text{min}}(f, k) = \frac{1}{|A(f)|^2} \quad (9)$$

که بعد از k بعد از فضای نویزی را نشان می دهد [۱].

۳-۳-۱- پیاده سازی روش مینیمم نرم

در روش مینیمم نرم هم ابتدا باید از معیار Minimum Description Length (MDL)، تعداد زیر فضای نویزی پیدا شود که معیار MDL به صورت زیر تعریف شده [۲]:

$$\begin{aligned} \text{MDL}(k) &= -N K \phi(k) \\ &+ \frac{1}{2} (m+1-k)(m+1+k) \log(N), k \\ &= 1, 2, \dots, m+1 \end{aligned} \quad (10)$$

که m ماکزیمم تعداد تاخیر در ماتریس اتوکوریلیشن (همان ترتیب بردارهای ویژه) را نشان می دهد، N تعداد نمونه های سیگنال، و $\phi(k)$ تابع likelihood است که به صورت مقابل تعریف می شود:

$$\phi(k) = \log \left[\frac{\prod_{i=0}^{k-1} (\lambda_i)^{\frac{1}{k}}}{\sum_{i=0}^{k-1} \frac{\lambda_i}{k}} \right] \quad (11)$$

و k تعداد زیر فضای نویزی است که با مینیمم کردن معیار MDL به دست می آید. همان طور که از فرمول ها پیداست، k از روش سعی و خطا به دست می آید.

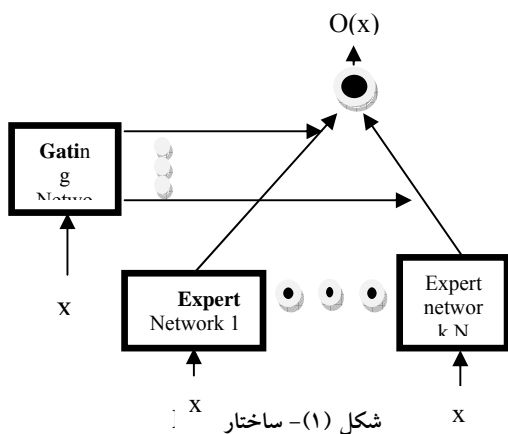
۴- بیان مدل شبکه عصبی

۴-۱- تجمیع خبرگان ME

ساختار ME نوشته شده با شبکه ورودی و چندین شبکه خبره در (شکل ۱) نشان داده شده است. شبکه ورودی بردار x را به عنوان ورودی می گیرد و خروجی های اسکالری تولید می کند که هر یک کسری از واحدند [۱].

هر یک از شبکه های خبره به ازای بردار ورودی یک بردار خروجی تولید می کند. شبکه ورودی تولید ضرایب ترکیب خطی برای ترکیب خروجی های شبکه های خبره را به عهده دارد، بنابراین، خروجی نهایی ساختار ME یک مجموع وزن دار همگرا از همه بردارهای خروجی تولید شده توسط شبکه های خبره است. فرض کنید تعداد N شبکه خبره در ساختار ME وجود دارد. شبکه های خبره به کار رفته همگی خطی هستند و تنها دارای یک تابع غیر خطی در خروجی هستند که با وجود این می توان از آن به عنوان "خطی تعمیم یافته" یاد کرد. شبکه خبره i ام که خروجی $O_i(x)$ خودش را به صورت یک تابع خطی تعمیم یافته از ورودی x تولید می کند، به این صورت تعریف می شود [۱]:

$$O_i(x) = f(W_i x) \quad (12)$$



که در آن W_i ماتریس وزن و $f(\cdot)$ یک تابع پیوسته غیر خطی مشخص است.

شبکه ورودی نیز یک تابع تعمیم یافته خطی است و خروجی i ام آن، $(g(x, v_i))$ ، یک چند جمله ای logit

یاتباع soft max از متغیرهای میانی Z_i است.

$$g(x, v_i) = \frac{e^{z_i}}{\sum_{k=1}^N e^{z_k}} \quad (13)$$

که در آن $z_i = v_i T x$ و v_i ها بردارهای وزن هستند.

خروجی کلی ساختار ME به شکل زیر است:

$$O(x) = \sum_{k=1}^N g(x, v_k) o_k \quad (14)$$

ساختار ME می‌تواند یک تفسیر احتمالاتی داشته

باشد. برای جفت ورودی-خروجی (x, y) ، مقادیر $g(x, v_i)$ را می‌توان به عنوان یک احتمال چند جمله‌ای فرض کرد که بیانگر احتمال تصمیمی باشد که با یک پروسه برگشتی x را به y نگاشت می‌کند. ابتدا باید تصمیم گرفته شود، که موجب انتخاب پروسه برگشتی i ام می‌شود، سپس خروجی y از چگالی احتمال $p(y-x, W_i)$ انتخاب می‌شود، همان ماتریس وزن از i امین شبکه خبره در مدل است، بنابراین، احتمال کلی ساختن y از ترکیبی از احتمالات ساختن y از هر جزء چگالی‌هاست [۱]:

$$P(y|x, \varphi) = \sum_{k=1}^N g(x, v_k) P(y|x, W_k) \quad (15)$$

که در آن φ دسته‌ای از همه پارامترهاست که شامل پارامترهای شبکه‌های خبره و شبکه ورودی است. بر پایه مدل احتمالی، آموزش در ساختار ME مثل مسأله \max likelihood رفتار می‌کند. Jordan و Jacob الگوریتم ماکزیمم کردن تابع EM را برای تعیین پارامترهای ساختار پیشنهاد می‌کنند. در این کار تعدادی شبکه خبره وابسته به هم می‌توانند با هم به شبکه ورودی برای تقسیم کردن مسأله طبقه‌بندی کلی به زیر مسأله‌های ساده‌تر استفاده شوند.

۴-۱-۱- پیاده سازی ساختار ME

۴-۱-۱-۱- ویژگی‌ها

در این مرحله ابتدا چهار ویژگی از هر یک از طیف‌های به دست آمده از هر سه روش را انتخاب شده‌اند. این چهار ویژگی عبارتند از: ماکزیمم، مینیمم، انحراف معیار و میانگین از طیف توان برای هر پنجره ۲۵۶ تایی از سیگنال. به این ترتیب دوازده ویژگی برای هر پنجره ۲۵۶ تایی پیدا می‌شود

که این دوازده ویژگی ورودی شبکه عصبی هستند [۱].

۴-۱-۱-۲ ساختار

این ساختار در واقع مجموعه‌ای از شبکه‌های عصبی است که وزن‌های این شبکه‌ها توسط قوانین احتمالی به روز می‌شوند [۴]. در این ساختار دو قسمت مهم دیده می‌شود که مربوط به دو سری از شبکه‌هاست. یک سری از شبکه‌ها در این ساختار خبره نامگذاری شدند و یک شبکه به نام شبکه ورودی وجود دارد [۶]. مطلب مهم دیگری که در این ساختار باید به آن اشاره کرد، ورودی شبکه‌های این ساختار است که به صورت ویژگی‌های مرکب است (در کار حاضر دوازده ویژگی که هر چهار ویژگی مربوط به یک روش خاص است). ابتدا یک سری شبکه (خبرگان) بردار ورودی را می‌گیرند و هر کدام یک بردار خروجی تولید می‌کنند. شبکه ورودی هم بردار ورودی را می‌گیرد و یک بردار خروجی تولید می‌کند که هر کدام از اجزای بردار خروجی شبکه ورودی به یکی از شبکه‌های خبره می‌رود. به این صورت خروجی شبکه‌های خبره وزن دار می‌شوند و در نهایت خروجی کلی سیستم، جمع وزن دار همه خروجی‌های شبکه‌های خبره است [۷-۱۵].

در مرجع [۱] ساختار استفاده شده برای شبکه‌های خبره و ورودی ۱۲-۲۰-۵ است، در حالی که در روش پیشنهادی پس از بررسی تعدادی از انواع ساختارهای معمول شبکه عصبی مصنوعی، بهترین شبکه متوجه که یک شبکه ورودی با ساختار ۱۲-۵۰-۵ و شبکه‌های خبره با ساختار ۱۲-۲۵-۵ است، انتخاب گردید. همچنین، تفاوت دیگر روش پیشنهادی نسبت به روش استفاده شده در مرجع ۱ تقسیم بندی تعداد مشاهدات برای دو مرحله آموزش و تست است؛ به این ترتیب که مرجع ۱ تعداد دادگان مساوی را برای مرحله آموزش و تست در نظر گرفته است، در حالی که در این مقاله با توجه به روش اعتبارسنجی^۷ برای انجام اعتبارسنجی، داده‌ها به ۷۰-۲۰-۱۰ تقسیم‌بندی شده‌اند؛ به این ترتیب که ۷۰٪ آنها برای آموزش، ۲۰٪ برای تست و ۱۰٪ برای اعتبارسنجی استفاده شده‌اند. این امر باعث می‌شود شبکه آموزش بهتری ببیند و همچنین، شبکه از حفظ کردن^۸ مصون می‌ماند. برای اجرای ساختار

فوق ابتدا باید:

برای هر جفت داده (منظور ورودی شبکه که خود دارای دوازه بعد و خروجی که پنج تایی است) احتمال پیشین را محاسبه کرده که نیازمند پیاده سازی فرمول (۱۶) است:

$$g(x, v_i) = \frac{e^{z_i}}{\sum_{k=1}^N e^{z_k}} \quad (16)$$

و همچنین باید احتمال شرطی $P(y|x, w_k)$ پیاده شود که به این منظور باید تابع توزیع احتمال موجود باشد. لذا در بررسی های انجام شده روی منابع ذکر شده تصمیم بر استفاده از احتمال چند جمله ای مطابق الگوی رفتاری به کار رفته در مراجع [۶-۸]، شده است.

۱- برای هر شبکه خبره مسأله IRLS که یک اپتیمم یاب است، پیاده شده (به این ترتیب وزن های update شده را می توان محاسبه کرد)، یاد آور می شود مسأله فوق برای هر شبکه با مشاهده y و x و وزن h باید پیاده شود [۹].

۲- برای شبکه gating (شبکه ورودی) مسأله IRLS با مشاهده x و h انجام شده است [۹].

۳- مراحل بالا تا حدی تکرار شود که به جواب مطلوب منجر شود [۹].

در نهایت، برای محاسبه خطا از حد آستانه با مقدار ۰/۰۵، استفاده شده است و شرط ورود به مرحله تست به دست آوردن حداقل صحت ۹۵٪ است. نتایج به دست آمده در جدول ۱ نشان داده شده است. شبکه MLP استفاده شده در شبکه ورودی دارای سه لایه پیشخور است که با استفاده از الگوریتم پس انتشار خطا با نرخ یادگیری متغیر آموزش داده می شود [۱۶، ۱۷].

تعداد نورون های لایه اول برابر با تعداد ویژگی ها (۱۲ نورون) در نظر گرفته شده و لایه خروجی شامل ۵ نورون است که مقادیری بین ۱ و ۰ اختیار می کنند. با تغییر تعداد نورون های لایه میانی، سعی در بهینه سازی معماری شبکه عصبی گردیده است. و سر انجام، یک شبکه عصبی سه لایه دارای ۵۰ نورون در لایه پنهان با تابع فعالیت سیگموئید استاندارد، انتخاب شده است. آموزش شبکه تا زمانی که مربع خطا کمتر از ۰،۰۱ شود یا تعداد تکرار های آموزش به ۱۰۰۰ برسد، ادامه می یابد. به همین ترتیب،

ساختار ۱۲-۲۵-۵ برای شبکه های خبره انتخاب شده است که بر قدرت بیشتر شبکه ورودی (که در واقع نقش انتخاب گری است) دلالت دارد. این مسأله به همراه تغییر تعداد داده های گروه های تست و آموزش (نسبت به [۱]) به منظور آموزش بهتر و همچنین، استفاده از اعتبارسنجی^۷ برای جلوگیری از حفظ کردن دادگان^۸، باعث ارتقای صحت طبقه بندی کننده شده است.

با توجه به نتایج به دست آمده به نظر می رسد ساختار پیشنهادی توانمندی به مراتب بیشتری در طبقه بندی سیگنال های EEG نسبت به ساختار ارائه شده در مرجع [۱] دارد.

با استفاده از تخمین فرکانس تولید شده به وسیله این روش ها سطوح توان سیگنال می توانند از روی ماتریس توان بیان شوند. یک پنجره مستطیلی ۲۵۶ تایی از داده برای ایستا فرض شدن سیگنال در وقفه ها استفاده شده است. برای هر قسمت، ۱۲۹ نقطه از لگاریتم سطح توان PSD محاسبه شد. ابعاد بزرگ بردار ویژگی، پیچیدگی محاسباتی را افزایش می داد و بنابراین، برای کاهش بعد بردار ویژگی استخراج شده (انتخاب ویژگی)، مشخصه هایی پیرامون PSD استفاده شد. ویژگی های آماری زیر برای کاهش بعد بردارهای ویژگی مذکور استفاده شده است: (۱) ماکزیمم سطوح توان در PSD به روش پیسارنکو برای هر قسمت از EEG، ماکزیمم سطوح توان در PSD به روش MUSIC برای هر قسمت از EEG، ماکزیمم سطوح توان در PSD به روش مینیمم نرم برای هر قسمت از EEG.

(۲) مینیمم سطوح توان در PSD به روش پیسارنکو برای هر قسمت از EEG، مینیمم سطوح توان در PSD به روش MUSIC برای هر قسمت از EEG، مینیمم سطوح توان در PSD به روش مینیمم نرم برای هر قسمت از EEG.

(۳) میانگین سطوح توان در PSD به روش پیسارنکو برای هر قسمت از EEG، مینیمم سطوح توان در PSD به روش MUSIC برای هر قسمت از EEG، مینیمم سطوح توان در PSD به روش مینیمم نرم برای هر قسمت از EEG.

۴) انحراف معیار سطوح توان در PSD به روش مینیمم سطوح توان در PSD به روش مینیمم نرم برای هر قسمت از EEG. مینیمم سطوح توان در PSD به روش MUSIC برای هر قسمت از EEG،

جدول (۱): نتایج صحت ساختار ME

نتایج بدست آمده از طبقه بندی کننده ME در ساختار پیشنهادی		میانگین (در ۱۰ بار اجرا شبکه)	واریانس (در ۱۰ بار اجرا شبکه)
درصد صحت آموزش	۹۹/۷۵۰۰	۹۹/۲۵۰۰	۹۹/۸۱۲۵
درصد صحت تست	۹۹/۲۸۵۷	۹۹/۳۷۵۰	۹۹/۳۹۲۸

۶- نتیجه گیری

در این مطالعه، ساختار ME برای کلاس بندی سیگنال EEG استفاده شده است. در این مطالعه ۱۰۰ سیگنال دارای ۴۰۹۶ نمونه برای هر کلاس با پنجره مستطیلی از ۲۵۶ داده ای تقسیم شدند و سپس داده ها به فرم ۸۰۰۰ بردار (۱۶۰۰ بردار برای هر کلاس) تبدیل شدند.

نکته قابل توجه این است که با وجود استفاده از اعتبار سنجی^۹ به منظور جلوگیری از حفظ کردن^{۱۰}، میزان خطای تست خیلی پایین است (یا می توان گفت صحت تست بالاست) که این حاکی از آموزش دیدن خوب شبکه در ضمن عدم حفظ کردن داده‌هاست. همان طور که ذکر شد، با تغییر ساختار شبکه‌ها در قسمت gating و خبره، قدرت شبکه در یادگیری داده ها افزایش یافته و از طرفی، زیاد شدن پارامتر های ناشی از افزایش قدرت شبکه (که امری نا مطلوب است)، با اختصاص دادن داده های بیشتر در فاز تعلیم به شبکه جبران شده و به صحت بالاتری در مقایسه با مرجع [۱] رسیده است (البته، نتایج فوق با اعمال حد آستانه ۰,۰۵ روی تفاضل خروجی نهایی از مقدار مطلوب شبکه به دست آمده است). اگرچه نباید از تاثیر منفی این امر (یعنی افزایش زمان رسیدن به جواب در مقایسه با مرجع [۱]) غافل بود، ولی بررسی نحوه تاثیر برآیند همه عوامل مثبت و منفی اشاره شده، نشان می دهد که استفاده از الگوریتم پیشنهادی توانمندی شبکه را نسبت به روش های ارائه شده در مقالات دیگر، بهبود داده است. ناگفته نماند که افزایش زمان رسیدن به جواب با توجه به پیچیدگی شبکه و تعداد

نورون‌های به کار رفته در ساختار شبکه، امری منطقی و قابل پیش بینی است. در جدول (۲) نتایج گزارش شده مرجع ۱ و این مقاله، در مقایسه با هم ارائه شده است. همچنین، نکته دیگری که با توجه به نتایج به دست آمده می توان اشاره کرد، این است که ترکیب چند شبکه عصبی حتی بدون استفاده از ساختار احتمالی در به دست آوردن وزن‌ها در مسأله طبقه بندی، می تواند تولید صحت بالاتر در زمان کمتر کند که این مسأله از نظر طبقه بندی مهم و ارزشمند است.

در مجموع، با توجه به نتایج [۱] و نتایج مقاله، روش پیشنهادی سیستم هوشمند ترکیبی نسبت به روش شبکه خبره کلاسیک (روش استفاده شده در مرجع ۱) برتری دارد، زیرا کاهش تعداد پارامترهای شبکه باعث ناتوانی شبکه است و از طرفی تعداد دادگان کم برای آموزش موجب عدم تعلیم کافی شبکه می شود و لازمه حل این مشکل، افزایش دادگان در فاز تعلیم است، به طوری که بتواند محدوده قابل قبولی از تنوعات را پوشش دهد. که در این مطالعه این موضوع با افزایش ۵۰٪ (برای مرجع [۱]) به ۷۰٪ محقق شده است.

جدول (۲): مقایسه نتایج ساختار پیشنهادی و ساختار بیان شده

در مقاله [۱]

زمان	صحت	روش
۱۴ دقیقه و ۵۱ ثانیه	۹۵/۵۳	ME در مقاله [۱]
۴۰ دقیقه و ۲۳ ثانیه	۹۹/۵۰	پیشنهادی ME

pp.473-493, 2003.

- [13] Hong, X., Harris, C.J., "A Mixture of Experts Network Structure Construction Algorithm for Modeling and Control", *Applied Intelligence*, Vol.16, No.1, pp.59-69, 2002.
- [14] Rao, B. D., Hari, K. V. S., "STATISTICAL PERFORMANCE ANALYSIS OF THE MINIMUM-NORM METHOD", *Radar and Signal Processing*, Vol.136, No.3, pp.2760-2763, 1989.
- [15] Chen, K., Xu, L., Chi, H., "Improved learning algorithms for mixture of experts in multiclass classification", *Neural Networks*, Vol.12, No.9, pp.1229-1252, 1999.
- [16] Ebrahimzadeh, E., Pooyan, M., "Early detection of sudden cardiac death by using classical linear techniques and time-frequency methods on electrocardiogram signals", *J. Biomedical Science and Engineering*, Vol.4, No.11, PP.699-706, 2011.
- [17] Ebrahimzadeh, E., Alavi, S, M. Bijar, A., Pakkhesal, A, R., "A novel approach for detection of deception using Smoothed Pseudo Wigner-Ville Distribution (SPWVD)" *J. Biomedical Science and Engineering*, Vol. 6, No.1, pp. 8-18, 2013.

زیر نویس ها

-
- ¹ Power Spectral Density
² Cross Validation
³ Epileptogenic
⁴ Hippocampal
⁵ Pisarenko
⁶ Spurious
⁷ Cross Validation
⁸ Over Fitting
⁹ Crossvalidation
¹⁰ Overfitting

مراجع

- [1] Ubeyli, E.D., Guler, I., "Features extracted by eigenvector methods for detecting variability of EEG signals", *Pattern Recognition Letters*, Vol.28, No.5, pp. 592-603, 2007.
- [2] Musselman M., Djurdjanovic D., "Time-frequency distributions in the classification of epilepsy from EEG signals ", *Expert Systems with Applications*, Vol.39, No.13, pp. 11413-11422, 2012.
- [3] Derya Übeyli E., "Wavelet/ mixture of experts network structure for EEG signals classification ", *Expert Systems with Applications*, Vol.34, No.3, pp. 1954-1962, 2008.
- [4] Jacobs, R. A., Jordan, M.I., "Adaptive Mixtures of Local- Experts", *Newel Computation*, Vol. 3, pp. 79-87, 1991.
- [5] Jong, W, Y., Sung, I, Y., Sung, B. C., "Adaptive mixture-of-experts models for data glove interface with multiple users", *Expert Systems with Applications*, Vol. 39, No.5, pp.4898-4907, 2012.
- [6] Jordan, M. I., Jacobs, R. A. "Hierarchical Mixtures of Experts and the EM Algorithm", *Neural Computation*, Vol.6, No.2, pp.181-214, 1993.
- [7] Subasi, A., "EEG signal classification using wavelet feature extraction and a mixture of expert model", *Expert Systems with Applications*, Vol. 32, No.4, pp. 1084-1093, 2007.
- [8] Chen, K., "A connectionist method for pattern classification with diverse features", *Pattern Recognition Letters*, Vol.19, No.7, pp. 545-558, 1998.
- [9] Ubeyli, E.D., Cvetkovic, D., Cosic, I., "Analysis of human PPG, ECG and EEG signals by eigenvector methods", *Digital signal processing letter*, Vol.20, No.3, pp.956-963, 2009.
- [10] Ebrahimpour,R., Kabir, E., Yousefi, M.R., "Teacher-directed learning in view-independent face with recognition mixture of experts using overlapping eigen spaces", *Computer Vision and Image Understanding* , Vol.111, No.2, pp. 195-206, 2008.
- [11] Jiang, W., Tanner, M.A., "On the identifiability of mixtures-of-experts", *NeuralNetworks*, Vol.12, No.9, pp.1253-1258, 1999.
- [12] Ubeyli, E.D., Guler, I., "Comparison of eigenvector methods with classical and model-based methods in analysis of internal carotid arterial Doppler signals", *Computers in Biology and Medicine*, Vol.33, No.6,