

Designing a Deep Fuzzy Rule-Based System for Depression Staging

R. Davoodi¹, M. H. Moradi^{2*}

¹Ph.D Student, Bioelectric Department, Biomedical Engineering Faculty, Amirkabir University of Technology, Tehran, Iran

² Professor, Bioelectric Department, Biomedical Engineering Faculty, Amirkabir University of Technology, Tehran, Iran

Receipt in the Online Submission System: 28 May 2018, Received in Revised Form: 22 July 2018, Accepted: 25 August 2018

Abstract

Depression is one of the most common mental disorders in the current century where early diagnosis can result in better treatment. One of the depression diagnostic methods is the analysis of the brain electrical signals. In this paper, we are seeking for a method to distinguish among the levels of the depression. The proposed model is a deep rule-based system based on the stacked principle and focuses on the interpretability of the rules alongside high accuracy. Fuzzy systems have the proper capability in the classification of medical data with various levels of uncertainty. Moreover, in the recent years, deep learning has been taken considerable attention in the field of Artificial Intelligence. In this paper, we aim to benefit from capabilities of both fields. The proposed architecture employs a robust fuzzy clustering approach that can determine an appropriate number of clusters in each layer, unsupervised and a hierarchical stacked structure to transfer the interpretable trained rules from the previous layers with the same linguistic labels to the next layer. The interpretability is due to the presence of the input space into the consequent ones. The presence of the output of the previous layer's rules at the input space of the next parts equals to a fuzzy system with non-linear consequent or the certainty factor in a fuzzy system with linear consequent. EEG data were preprocessed and time, frequency and nonlinear features such as recurrent plot were extracted and selected and after that were employed in the proposed system. The proposed system was compared with common classifiers like Neural Net, Support Vector Machine, Naive Bayes, Decision Tree and Linear Discriminant Analysis. Accuracy results for the test data in 30 folds (49.01% in comparison to 41.42%, 40.47%, 40.01%, 38.35% and 40.28% respectively) demonstrate the considerable performance of the proposed system.

Keywords: *Classification, Depression, Fuzzy Network, Deep learning*

*Corresponding Author

Address: Bioelectric Department, Biomedical Engineering Faculty, Amirkabir University of Technology, P.O. Box: 11365-9567, Tehran, Iran
Tel: +98-21-64542399
Fax: +98-21-66468186
E-mail: mhmoradi@aut.ac.ir

طراحی یک سیستم فازی عمیق مبتنی بر قاعده به منظور تعیین سطح افسردگی

راحله داودی^۱، محمدحسن مرادی^{۲*}

^۱ دانشجوی دکتری مهندسی پزشکی، گروه بیوالکترونیک، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، تهران
^۲ استاد، گروه بیوالکترونیک، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، تهران

تاریخ ثبت در سامانه: ۱۳۹۷/۳/۷، بازنگری: ۱۳۹۷/۴/۳۱، پذیرش قطعی: ۱۳۹۷/۶/۳

چکیده

افسردگی یکی از شایع‌ترین اختلالات روانی عصر حاضر است که تشخیص زودهنگام شدت آن می‌تواند در روند درمان مفید باشد. یکی از روش‌های تشخیص این بیماری، تحلیل اطلاعات حاصل از سیگنال‌های الکتریکی مغزی می‌باشد. در این مقاله، به دنبال تمایز میان سطوح افسردگی با استفاده از تحلیل سیگنال مغزی هستیم. مدل پیشنهادی، سیستم عمیق مبتنی بر قاعده با استفاده از قابلیت پشته است و تمرکز روی تفسیرپذیری قواعد در کنار دقت بالا می‌باشد. سیستم‌های فازی قابلیت مناسبی را در طبقه‌بندی دادگان پزشکی با عدم قطعیت نشان داده‌اند. افزون بر این، در سال‌های اخیر یادگیری عمیق، توجه ویژه‌ای را در حوزه‌ی هوش مصنوعی کسب کرده است. در این مقاله به دنبال بهره‌گیری از قابلیت‌های هر دو رویکرد، در قالب یک سیستم فازی عمیق هستیم. سیستم پیشنهادی از یک رویکرد خوشه‌بندی مقاوم بهره می‌برد که قادر است تعداد خوشه‌های بهینه برای هر لایه را به صورت بدون سرپرست تعیین نماید. در کنار آن، مدل پیشنهادی از یک ساختار سلسله‌مراتبی پشته‌ای بهره می‌برد، به این صورت که قواعد آموزش‌یافته‌ی تفسیرپذیر در لایه‌ی اول را با برچسب‌های زبانی یک‌سان برای تمام ورودی‌ها، به صورت خروجی لایه‌ی اول در کنار ورودی، به لایه‌ی بعد منتقل نماید. وجود خروجی قواعد لایه‌های قبل در فضای ورودی لایه‌های بعد معادل قابلیت اطمینان در سیستم فازی با تالی خطی یا یک سیستم فازی با تالی غیر خطی می‌باشد. دادگان مورد استفاده پس از پیش‌پردازش، استخراج ویژگی‌های زمانی، فرکانسی و غیرخطی نظیر بعد نگاشت بازگشتی و کاهش بعد، به سیستم پیشنهادی ارائه شد. سیستم پیشنهادی با طبقه‌بندهای متداول نظیر شبکه‌ی عصبی، ماشین بردار پشتیبان، بیز ساده، درخت تصمیم و آنالیز افتراقی خطی مقایسه شد. نتایج صحت دادگان تست به دست آمده در ۳۰ تکرار (۴۹،۰۱٪) در مقابل به ترتیب ۴۱/۳۲٪، ۴۰/۴۷٪، ۴۰/۰۱٪، ۳۸/۳۸٪ و ۴۰/۲۸٪، بیان‌گر قابلیت قابل توجه این مدل در تفکیک چهار سطح افسردگی می‌باشد.

کلیدواژه‌ها: طبقه‌بندی، افسردگی، شبکه‌ی فازی، یادگیری عمیق

*نویسنده مسئول

نشانی: گروه بیوالکترونیک، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، تهران، ایران، کد پستی: ۱۲۳۴۵۶۷۸۹۰

تلفن: ۶۴۵۴۲۳۹۹ (۲۱) ۹۸+

دورنگار: ۶۶۴۶۸۱۸۶ (۲۱) ۹۸+

پست الکترونیکی: mhmoradi@aut.ac.ir

۱- مقدمه

افسردگی یکی از شایع‌ترین انواع اختلالات روانی است که موجب مراجعات مکرر افراد به مراکز درمانی و متخصصان امر می‌شود. در سطح جهانی، حدود ۳۰۰ میلیون انسان از افسردگی رنج می‌برند که معادل ۴/۴٪ از کل جمعیت جهان می‌باشد [۱]. روش اصلی برای تشخیص این اختلال، پرسش و پاسخ شفاهی بین متخصص و بیمار و پرسش‌نامه‌های کتبی است که بر اساس معیارهای استاندارد برای تشخیص این اختلال طراحی شده‌اند. متخصصان حداقل یکی از ابزارهای مبتنی بر گزارش شخصی بیمار را به منظور ارزیابی درجه‌ی افسردگی به هنگام اولین مراجعه‌ی بیمار مورد استفاده قرار می‌دهند و از نتایج آن برای پیگیری تغییرات حاصله در مقدار افسردگی در طول مدت درمان بهره می‌گیرند، که یکی از این پرسش‌نامه‌ها، پرسش‌نامه‌ی افسردگی بک^۱ می‌باشد. این پرسش‌نامه در آغاز هر جلسه، توسط بیمار تکمیل می‌گردد و بر اساس این شاخص، درجه‌ی افسردگی در سطوح خفیف، ملایم، متوسط و شدید تعیین می‌شود.

یکی از روش‌هایی که برای تشخیص اختلال افسردگی مورد استفاده قرار می‌گیرد، تحلیل سیگنال الکتریکی مغزی^۲ است که به دلیل ثبت غیرتهاجمی، دقت زمانی مناسب، سهولت ثبت، هزینه‌ی کم و امکان تحرک و آزادی عمل برای انجام فعالیت حین ثبت، از اهمیت ویژه‌ی برخوردار است [۲، ۳، ۴].

سیگنال EEG به دلیل ماهیت بیولوژیکی و نحوه‌ی ثبت آن‌ها، از انواع ابهام و پیچیدگی برخوردار هستند. استفاده از رویکردهایی که قادر به مدل‌سازی این پیچیدگی‌ها باشد، از اهمیت ویژه‌ی برخوردار است. رویکردهای مختلفی برای مدیریت پیچیدگی و ابهام در استخراج ویژگی، مدل‌سازی و طبقه‌بندی آن وجود دارد که از جمله می‌توان به استخراج ویژگی‌های غیرخطی و آشوب‌ناک و به کارگیری طبقه‌بندهای غیرخطی و چندلایه، نظیر سیستم‌های فازی و مدل‌های عمیق اشاره کرد. سیستم‌های فازی نقشی اساسی در مواجهه با عدم قطعیت در تصمیم‌گیری فضای پزشکی ایفا می‌کنند.

از سوی دیگر، یادگیری عمیق، ابزار قدرتمندی برای حل مسایل طبقه‌بندی پیچیده است و در کاربردهای مختلف نظیر بینایی ماشین، پردازش زبان طبیعی، آنالیز ویدئو و فیزیک استفاده

شده است. با این وجود، اخیراً محققان اقدام به معرفی این تکنولوژی در حوزه‌های علوم شناختی و واسط مغز-رایانه نموده‌اند. شبکه‌های عمیق و چند لایه به عنوان ابزاری محاسباتی برای تشخیص بیماری‌ها شناخته شده‌اند که به دلیل شباهت به ساختار چند لایه‌ی مغز انسان، قدرت تمایز بیش‌تری نسبت به روش‌های تحلیل سیگنال دارند [۵].

ساختارهای عمیق و منطق فازی، هر کدام با رویکردهای خاص خود، به دنبال مدل‌سازی و حل مسایل پیچیده هستند. به جهت بهره بردن از قابلیت‌ها و نقاط قوت هر دو رویکرد فازی و عمیق، دسته‌ی قابل توجهی از مقالات هستند که اخیراً به سمت ترکیب رویکردهای محاسبات نرم رفته‌اند، که در ادامه به آن‌ها اشاره خواهد شد.

هدف این مقاله، ارائه‌ی یک ساختار فازی عمیق است که به صورت یک طبقه‌بند چندکلاسه معرفی می‌شود. نوآوری دیگر این مقاله، سطح‌بندی شدت افسردگی در سیگنال EEG است که با استفاده از ویژگی‌های غیرخطی مستخرج از این سیگنال، نیازمند طبقه‌بندی پیچیده و قوی می‌باشد و به دلیل ماهیت غیرخطی و غیرایستای EEG، کاربرد مناسبی برای بهره‌گیری از طبقه‌بند پیشنهادی است. از موارد قابل توجه در یادگیری عمیق، حجم دادگان مناسب جهت آموزش تعداد قابل توجه مجهولات مدل می‌باشد که در این مقاله تلاش شده است تا با استخراج تعداد مناسب دنباله‌های سیگنال EEG (حدود ۱۴۰۰ نمونه) و به کارگیری روش مناسبی برای تعلیم بخشی از مجهولات، با این چالش مواجهه نماید.

ساختار این مقاله بدین صورت است که در ادامه مروری بر مطالعات انجام شده در دو حوزه‌ی کاربرد طبقه‌بندی افسردگی و روش‌های ترکیب سیستم‌های فازی و یادگیری عمیق صورت می‌گیرد و در بخش دوم، دادگان مورد استفاده معرفی و روش پیشنهادی به تفکیک شرح داده می‌شود. در ادامه، نتایج به دست آمده به صورت مبسوط گزارش شده و مورد بحث قرار داده می‌شود و در نهایت جمع‌بندی صورت می‌گیرد.

۲- مروری بر مطالعات صورت‌گرفته

با توجه به بدیع بودن کاربرد طبقه‌بندی شدت افسردگی با استفاده از سیگنال EEG، نمی‌توان در این خصوص به پژوهشی

^۲ Electro Encephalo Graphy (EEG)^۱ Beck Depression Inventory (BDI)

تکنیک‌های پیش‌تاز در هوش مصنوعی در دهه‌ی گذشته قلمداد می‌شود. از این روی، بهره بردن از نقاط قوت هر دو رویکرد، مورد توجه بسیاری از مطالعات بوده است. ترکیب شبکه‌های عمیق و سیستم‌های فازی، از راه‌های مختلف ممکن می‌باشد. یک راه، بهینه‌سازی شبکه‌های عصبی عمیق^۱ نشات گرفته از محاسبات نرم است. این تکنیک‌ها، به همراه دیگر روش‌های محاسبات نرم، برای بهینه‌سازی DNN استفاده می‌شوند که عمل‌کرد یادگیری عمیق را بهینه می‌سازند. راه‌های دیگر که در آن‌ها تکنیک‌های یادگیری نرم توسط قابلیت استخراج ویژگی DNN حمایت می‌شوند، عموماً در کاربردهای خاص‌تر برای نتیجه‌ی بهتر استفاده می‌شوند. روش پیشنهادی در این مقاله به دسته‌ی دوم نزدیک‌تر است.

مجموعه‌های فازی و DBN توسط ژو و هم‌کارانش در [۱۲] به کار گرفته شد تا مفاهیمی از یادگیری نیمه‌نظارتی را برای طبقه‌بندی جملات استفاده نمایند. در این سیستم که شبکه‌های باور عمیق فازی^{۱۱} نام‌گذاری شد، با استفاده از یک ساختار عمیق آموزش‌یافته، توابع عضویت فازی برای هر کلاس از مرورها طراحی، و توزیع فضا برای تمام مشاهدات به عنوان دانش اولیه در نظر گرفته شد. توسط تعدادی از توابع عضویت فازی کدگذاری و سپس بر مبنای شبکه‌ی به دست آمده و توابع فازی، ساختار FDBN ساخته شد.

ساختار دیگری از شبکه‌ی فازی عمیق، توسط ژانگ [۱۳] ارائه شد که در آن ماشین بولتزمن محدود عمیق فازی فیثاغورثی^{۱۲} طراحی شد. پارامترهای این مدل، اعداد فازی فیثاغورثی است و هر نورون قادر خواهد بود تا تاثیر هر ویژگی را در تولید خروجی صحیح از هر دو جنبه‌ی مثبت و منفی یاد بگیرد. PFDBN توسط الگوریتم ترکیبی مبتنی بر گرادیان و روش تکاملی تعلیم یافت.

شبکه‌ی عصبی عمیق فازی با ساختار سلسله‌مراتبی ذوب شده^{۱۳} توسط دنگ و هم‌کارانش [۱۴] معرفی گشت که در آن مفاهیم یادگیری فازی در یادگیری عمیق معرفی شد. چندین قاعده‌ی

اشاره کرد. بنابراین نزدیک‌ترین حوزه به این موضوع، تشخیص بیماران افسرده از سالم است که در ادامه به صورت مختصر مرور می‌شود. در بخش ۲-۲ نیز به بررسی مجموعه‌ی مطالعاتی پرداخته می‌شود که ترکیبی از ساختارهای فازی و عمیق را ارائه نموده‌اند.

۲-۱- مطالعات انجام‌گرفته در تحلیل سیگنال مغزی برای تشخیص شدت افسردگی

بیش‌ترین مطالعات روی سیگنال مغزی برای بررسی افسردگی، بر اساس روش‌های خطی انجام شده و تعداد کمی از آن‌ها از تحلیل‌های غیرخطی بهره برده‌اند. بعضی از تحقیقات، از روش‌های آماری نظیر تحلیل واریانس^۱ و آزمون تی^۲ بهره برده‌اند [۶] و در برخی دیگر، از ویژگی‌های خطی و غیرخطی نظیر ویژگی‌های مولفه‌های زمانی، آماری، طیف توان در باندهای فرکانسی EEG، مولفه‌های زمان-فرکانسی و ویولت، تحلیل مولفه‌های مستقل^۳ و تصاویر sLORETA در کنار انواع ویژگی‌های آشوبی نظیر نمای لیاپانوف، بعد همبستگی، تحلیل نوسانات کاهش‌یافته^۴، فرکتال هیگوجی^۵ و غیره در کنار طبقه‌بندی‌های متداول نظیر ماشین بردار پشتیبان^۶ [۷]، طبقه‌بند خطی^۷ [۸]، شبکه‌ی عصبی^۸ [۹] و شبکه‌های عمیق^۹ [۱۰] تفکیک دو یا چند گروه صورت گرفته است. با این وجود، درباره‌ی جداسازی سطوح مختلف افسردگی، تاکنون گزارش قابل توجهی ارائه نشده است.

۲-۲- ترکیب سیستم‌های فازی با یادگیری عمیق

در [۱۱]، مروری بر تکنیک‌های محاسباتی نرم مبتنی بر یادگیری عمیق در کاربردهای مختلف صورت گرفته است. محاسبات نرم به دسته‌ای از روش‌ها گفته می‌شود که در ساخت نسل جدید از سیستم‌های محاسباتی هوشمند اهمیت دارند و موفقیت بسیار چشم‌گیری در حل مسایل محاسباتی عملی کسب کرده‌اند. از سوی دیگر، یادگیری عمیق از جمله

^۱ Neural Network (NN)

^۲ Deep Networks

^۳ Deep Neural Network (DNN)

^۴ Fuzzy Deep Belief Network (FDBN)

^۵ Pythagorean Fuzzy Deep Boltzmann Machine (PFDBM)

^۶ Hierarchical Fused Fuzzy Deep Neural Network (HFDNN)

^۱ ANOVA

^۲ T-test

^۳ Independent Component Analysis (ICA)

^۴ Detrended Fluctuation Analysis (DFA)

^۵ Higuchi fractal dimension

^۶ Support Vector Machines (SVM)

^۷ Linear Discriminant Analysis (LDA)

۳-۱- مشخصات داده

این مقاله از دادگان یک پژوهش توصیفی مقطعی استفاده می‌کند که به منظور تفکیک شدت افسردگی در چهار سطح بر اساس تحلیل سیگنال EEG می‌باشد. به این منظور، ۶۸ بیمار مبتلا به اختلال افسردگی با سطوح مختلف در نظر گرفته شدند که سطح‌بندی آن‌ها بر اساس خودارزیابی پرسش‌نامه‌ی بک افسردگی ضعیف (نمره‌ی بین ۱۱ تا ۱۶)، افسردگی ملایم (نمره‌ی بین ۱۷ تا ۲۰)، افسردگی متوسط (نمره‌ی بین ۲۱ تا ۳۰) و افسردگی شدید (نمره‌ی بالای ۳۰) صورت گرفت [۱۹]. بر این اساس، ۱۶ نفر افسردگی ضعیف، ۱۱ نفر افسردگی ملایم، ۱۰ نفر افسردگی متوسط و ۳۱ نفر افسردگی شدید برجسب گذاری شدند. سیگنال EEG در حالت استراحت و چشمان بسته با ۱۹ کانال مطابق با استاندارد ۱۰/۲۰ در کلینیک تخصصی مرکز بیماری‌های اعصاب و روان آتیه ثبت شد [۲۰]. پیش‌پردازش ابتدایی شامل فیلتر فرکانسی ۰/۱ تا ۴۰ هرتز و حذف آرتیفکت‌های چشمی می‌باشد. سیگنال هر فرد به قطعاتی ده ثانیه‌ای تقطیع شده (در مجموع ۱۳۸۵ قطعه‌ی سالم) و از هر قطعه، ۲۶ ویژگی خطی (زمانی و فرکانسی) و غیرخطی شامل توان زیرباند (باندهای آلفا، بتا، تتا و گاما)، کورتوسیس^۴، اسکینوس^۵، آنتروپی فازی^۶، بعد فرکتال فازی^۷، بعد فرکتال هیگوچی و ویژگی‌های بازگشتی^۸ استخراج گردید.

ویژگی‌های بازگشتی توسط اکمن و هم‌کارانش [۲۱] معرفی شد، و در بسیاری از مطالعات برای نمایش جزئیات سیگنال غیرایستای EEG در زمان‌های مختلف استفاده شده است [۱۹، ۲۰]. ایده‌ی اصلی این روش در بازگشت‌های حالت سیستم به مناطق یک‌سانی از فضای فاز، که قبلاً مسیر در آن‌ها بوده است، می‌باشد. تکرار یک رخداد در دو زمان به صورت ماتریسی مربعی از صفر و یک مشخص می‌شود. با انتخاب ویژگی مناسب از ماتریس بازگشتی، امکان آشکارسازی الگوهای مخفی و تغییرات ساختاری در دینامیک فراهم می‌شود. این قابلیت به عنوان آنالیز کمی بازگشتی شناخته می‌شود و از این ماتریس، ویژگی‌های زیر قابل استخراج است که در جدول (۱) به آن‌ها اشاره شده است.

منطقی در یادگیری فازی برای کاهش عدم قطعیت استفاده شد و استخراج اطلاعات به صورت سلسله‌مراتب در یادگیری عمیق به جهت کاهش نویز به کار گرفته شد که از آن، ویژگی‌های یادگیری فازی و عمیق استخراج شدند. توان‌مندی این مدل از هم‌جوشی اطلاعات فازی و عصبی در کنار هم ناشی شده است. ساختار دیگری از ترکیب از DNN برای کاربردهای تجاری معرفی شد که محدوده‌ی نوسانات در نرخ مبادلات را بر اساس سری‌های زمانی مبتنی بر مقادیر فازی و DBN پیش‌بینی می‌کند [۱۵]. در مقاله‌ای دیگر [۱۶]، یک معماری فازی یادگیری عمیق طراحی شد که در آن در لایه‌ی اول، توابع عضویت، در لایه‌ی دوم یک اپراتور نرم T و در لایه‌ی سوم یک تابع رگرسیون خطی قرار داشت. در لایه‌ی آخر، خروجی بر اساس قواعد اگر-آن‌گاه فازی تعیین می‌گردید. در [۱۷]، بر مبنای اصل تعمیم‌پذیری پشت‌های، یک طبقه‌بند فازی SKT^۱ معرفی شد که به دنبال یافتن قواعد فازی تفسیرپذیر بود. مشابه یادگیری‌های عمیق متداول، D-TSK-FC دارای ساختار لایه به لایه بود که با استفاده از ماشین یادگیری کمینه^۲، وزن‌های هر لایه آموزش داده می‌شد. فضای ورودی تمام لایه‌های طبقه‌بند از ترکیب فضای ورودی با جابه‌جایی خطی ماتریس پروجکشن در خروجی لایه‌ی قبل به دست آمده بود. در نتیجه، واحد پنهان در هر لایه از این ساختار بر قواعد فازی تفسیرپذیر مبتنی هستند.

با استفاده از رویکرد تفسیرپذیری، یک ساختار عمیق مبتنی بر قواعد فازی در [۱۸] معرفی شد که فضای ورودی با استفاده از یک خوشه‌بندی نرم به فضای قواعد نگاشت می‌شد و سپس با استفاده از ساختار چند لایه و قواعد تفسیرپذیر، ورودی هر لایه به صورت ترکیبی خطی از ورودی و خروجی لایه‌ی قبل ساخته می‌شد. در نهایت، خروجی نهایی سیستم، با استفاده از روش ترکیب رای غالب وزن‌دار^۳ تعیین می‌شد.

۳- مواد و روش‌ها

در این بخش مشخصات داده‌ی مورد استفاده و ساختار سیستم پیشنهاد شده به تفصیل شرح داده خواهد شد.

^۱ Skewness

^۲ Fuzzy entropy

^۳ Fuzzy fractal

^۴ Recurrence quantification analysis (RQA)

^۱ Deep TSK Fuzzy Classifier (D-TSK-FC)

^۲ Least Learning Machine (LLM)

^۳ Weighted Majority Voting

^۴ Kurtosis

جدول (۱) - معرفی ویژگی‌های بازگشتی استخراج شده

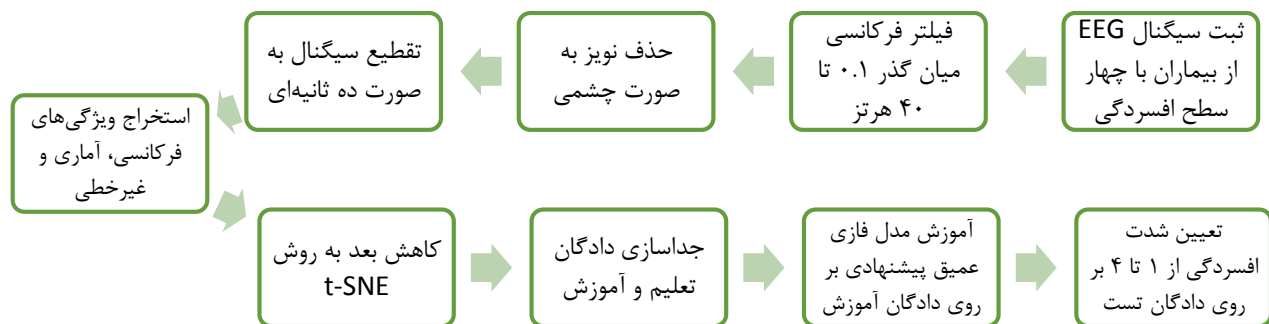
از سیگنال EEG

شرح ویژگی	ویژگی
Determinism (DET)	قطعییت
Recurrence Rate (RR)	نرخ بازگشت
Mean diagonal line length (L_{mean})	میانگین طول خطوط قطری
Longest diagonal line	طولانی‌ترین خط قطری
Entropy	آنتروپی
Laminarity (LAM)	لایه‌ای بودن
Trapping time (TT)	زمان به دام افتادن
Longest vertical line (V_{max})	طولانی‌ترین خط عمودی
Recurrence times T_1	زمان بازگشت نوع ۱
Recurrence times T_2	زمان بازگشت نوع ۲
Recurrence times entropy (RTE)	آنتروپی زمان بازگشت
Transportability	انتقال پذیری

برای اطلاعات بیشتر در خصوص جزئیات RQA به [۲۴] مراجعه شود.

در مجموع، ویژگی‌های به دست آمده با استفاده از روش کاهش بعد t-SNE [۲۵] به ۱۰ بعد محدود گردید. رویکرد t-SNE یک رویکرد مبتنی بر جایزه به برنده است. این روش با استفاده از نسخه‌ای از روش جاسازی تصادفی همسایه^۲ تلاش می‌نماید تا نقاط نمونه را در مرکز نقشه قرار دهد و برای کاهش قابل توجه ابعاد فضای ورودی استفاده می‌شود.

در گام یادگیری از ۷۰ درصد دادگان برای آموزش و ۳۰ درصد باقی‌مانده برای ارزیابی مدل استفاده شد و نتایج به ازای ۳۰ بار تکرار به صورت متوسط و انحراف معیار گزارش شده است. در شکل (۱) روند کلی مراحل صورت گرفته برای سطح‌بندی افسردگی در این مقاله نشان داده شده است.



شکل (۱) - بلوک دیاگرام روند مقاله

در بخش مقدم قواعد، آموزش با استفاده از روشی مبتنی بر خوشه‌بندی فازی c میانگین^۴ صورت می‌گیرد. در خوشه‌بندی فازی، الگوریتم FCM پراستفاده‌ترین الگوریتمی است که به کار می‌رود. اما FCM و تمام توسعه‌های مختلف آن تحت تاثیر مقاردهی اولیه به پارامترهایشان قرار دارند، که از آن جمله می‌توانیم به تعیین تعداد کلاسترها اشاره نماییم که باید پیش از اعمال الگوریتم FCM مشخص شده باشند. اگرچه روش‌هایی برای حل این مساله در FCM پیشنهاد شده، اما هیچ‌کدام به صورت هم‌زمان با راه حل‌های ارائه شده، به اندازه‌ی کافی قوی نبوده‌اند که هم مقاردهی اولیه انجام دهند و هم انتخاب

۳-۲- سیستم فازی عمیق مبتنی بر قواعد

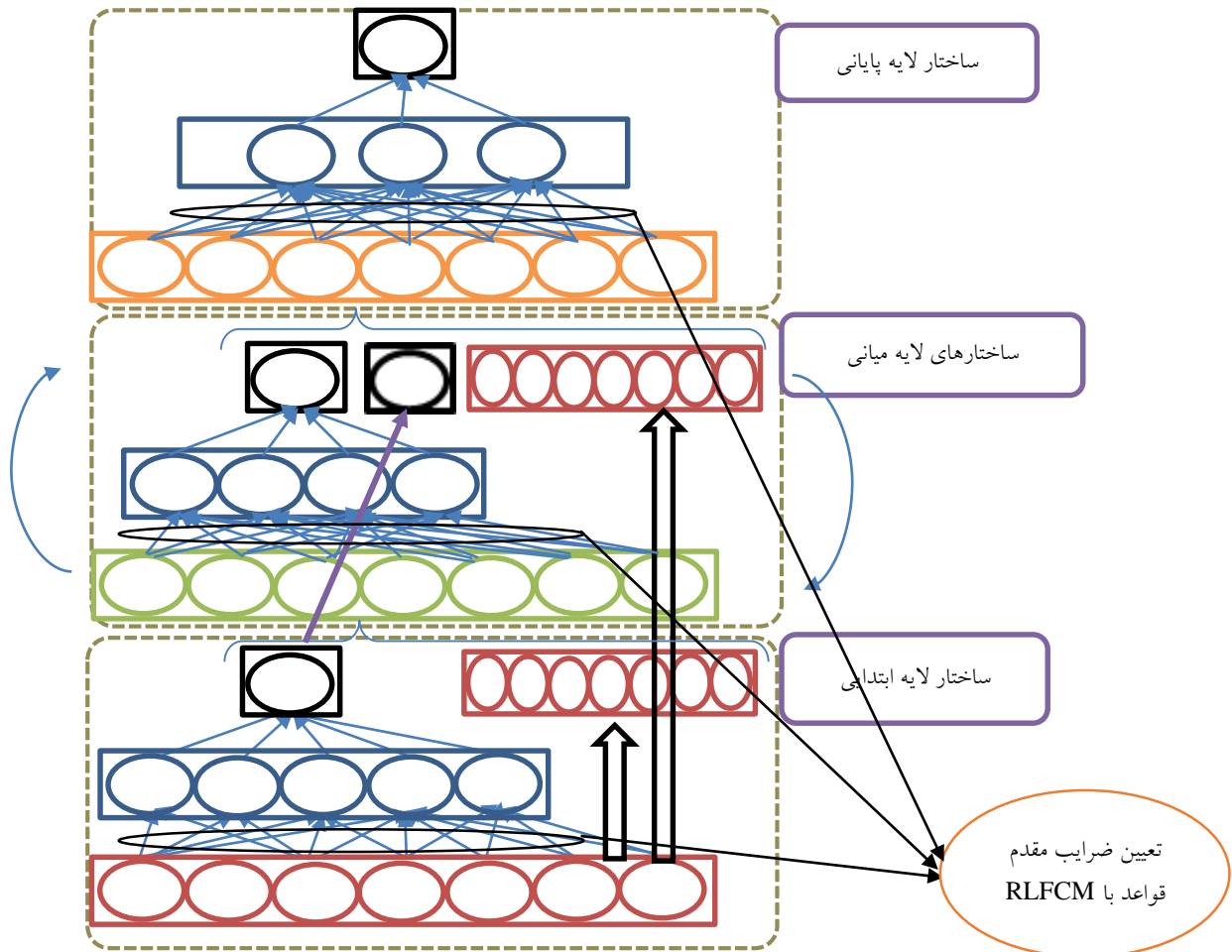
در این مقاله، یک معماری از سیستم فازی عمیق مبتنی بر قواعد، ارائه شده است. در بسیاری از کاربردهای طبقه‌بندی، نه تنها دقت بالا، بلکه قابلیت تفسیرپذیری نیز مورد نیاز است. در میان انواع طبقه‌بندی‌های موجود، طبقه‌بند فازی تاکاگی-سوگینو-کانگ^۳ یکی از بهترین گزینه‌ها بوده که قادر است این موازنه را میان دقت و تفسیرپذیری برقرار سازد. به همین جهت، در این ساختار از یک مدل TSK مرتبه‌ی اول با ساختار چند لایه و انتقالی به صورت پشته استفاده شده است.

^۲ Takagi-Sugeno-Kang (TSK)^۴ Fuzzy C-means (FCM)^۱ t-Distributed Stochastic Neighbor Embedding^۲ Stochastic Neighbor Embedding

اندیس فازی m ، اطمینان بخشی در عدم انتخاب پارامترها و اطمینان بخشی در عدم انتخاب تعداد خوشه‌ها قابلیت خود را نشان می‌دهد.

در ادامه ابتدا ساختار کلی مدل پیشنهادی مطرح می‌گردد، سپس نحوه‌ی عمل‌کرد الگوریتم RL-FCM، نحوه‌ی آموزش شبکه‌ی کلی و در نهایت فرایند ارزیابی شبکه ارائه می‌شود.

پارامتر، بی آن‌که از شاخص فازی بدون تعداد خوشه‌ای معین استفاده نمایند. در [۲۶] روش خوشه‌بندی فازی با یادگیری مقاوم^۱ ارائه شده که باعث می‌شود FCM بدون انتخاب اندیس فازی m و مقداردهی‌های اولیه و انتخاب پارامترها، به صورت خودکار بهترین تعداد خوشه‌ها را مشخص نماید. در واقع الگوریتم RL-FCM در اطمینان بخشی در مقداردهی اولیه بدون



شکل (۲) - ساختار مدل فازی عمیق پیشنهادی

و خروجی قواعد فازی به عنوان بلوک پایانی در هر لایه در نظر گرفته می‌شود. اتصال میان لایه‌های مختلف شبکه به صورت پشته است، به این صورت که نقاط ورودی به همراه خروجی به دست آمده از لایه‌ی قبل به عنوان بلوک ورودی لایه‌ی بعد در نظر گرفته می‌شود و تصمیم‌گیری در لایه‌ی فوق انجام می‌شود. در هر لایه از شبکه‌ی پیشنهادی، از قواعد فازی TSK استفاده

۳-۲-۱- ساختار مدل

در این ساختار از یک مدل TSK مرتبه‌ی اول با ساختار چند لایه و انتقالی با استفاده از قاعده‌ی تعمیم‌یافته‌ی پشته‌ای^۲ استفاده شده است. در هر لایه از ساختار عمیق دادگان به عنوان بلوک ورودی در نظر گرفته می‌شوند. قواعد فازی که با استفاده از خوشه‌بندی مقاوم به دست آمده‌اند، بلوک میانی را می‌سازند

^۲ Stacked generalization principle

^۱ Robust Learning Fuzzy C-means (RL-FCM)

ایده‌ی اصلی روش ارائه شده در [۲۶] به این صورت است که نسبت به انتخاب پارامترهای اولیه و انتخاب تعداد خوشه‌ها مقاوم باشد، بنابراین در تکرار اول، تمام داده‌ها را به عنوان مرکز خوشه در نظر می‌گیرد و بعد در تکرارهای بعدی با الگوریتمی به تعداد کلاسترهای واقعی موجود در داده‌ها هم‌گرا می‌شود.

به علاوه، پارامتر α که در هر خوشه‌ی k به عنوان نسبت ترکیب شدن در کلاستر در نظر گرفته می‌شود، به نحوی وزن هر خوشه را تعیین می‌کند و در واقع میزان عضویت هر داده به یک خوشه توسط آن مشخص می‌شود. پارامتر α جایگزین ضریب فازی بودن m در تعیین فاصله می‌باشد و خوشه‌هایی که برای آن‌ها این ضریب کم‌تر از مقدار مشخصی باشد، به عنوان خوشه‌های نامناسب، حذف می‌شوند. استفاده از آنتروپی درجه‌ی تعلق فازی به جای تعلق فازی، به این معنی است که α_k احتمال آن است که یک داده به خوشه‌ی k -ام متعلق باشد. با این حساب $\ln \alpha_k$ معادل اطلاعات حضور آن داده در آن خوشه است. با استفاده از این مفهوم، از روش یادگیری EM برای تنظیم مقدار α_k استفاده می‌شود. در نتیجه، عبارات آنتروپی موجود در الگوریتم، متناسب با ترکیب احتمالاتی در عضویت‌های فازی برای یادگیری بهترین تعداد خوشه‌ها استفاده می‌شود.

جزئیات بیش‌تر درباره‌ی تابع هزینه، در [۲۶] بیان شده است. برای کمینه‌سازی تابع هدف، مقادیر مجهول تابع عضویت، مراکز خوشه‌ها و مقدار α به صورت بیان شده در الگوریتم ۱ به دست می‌آیند.

الگوریتم ۱: روش خوشه‌بندی اولیه‌ی فازی

ورودی: داده‌ی آموزش

$$x_i \in \mathbb{R}^d \quad X_{train} = \{x_1, x_2, x_3, \dots, x_N\}^T$$

حداکثر تعداد تکرار و مقدار آستانه‌ی ϵ .

خروجی: مراکز خوشه‌ها $\mathbf{v}_k^{(t)}$ و α_k و c

مقداردهی اولیه:

$$\begin{aligned} c^{(0)} &= n \\ \mathbf{v}_k^{(0)} &= x_i \\ \mathbf{a}_k^{(0)} &= 1/n \\ \mathbf{r}_1^{(0)} = \mathbf{r}_2^{(0)} = \mathbf{r}_3^{(0)} &= \mathbf{1} \\ t &= 1 \\ \epsilon &> 0 \end{aligned}$$

تکرار حلقه تا شرط توقف:

۱- با استفاده از مقادیر $\mathbf{v}_k^{(t-1)}$ ، $\mathbf{a}_k^{(t-1)}$ ، $\mathbf{r}_1^{(t-1)}$ ، $\mathbf{r}_2^{(t-1)}$ و معادله‌ی زیر مقدار $\mu_{ik}^{(t)}$ برای تمام نمونه‌های مشاهده و تمام خوشه‌ها محاسبه شود.

می‌شود. در بخش مقدم قواعد از نسخه‌ی بهبودیافته‌ی الگوریتم خوشه‌بندی FCM استفاده می‌شود که در بخش بعد توضیحی اجمالی درباره‌ی آن ارائه می‌شود. در بخش تالی قواعد، با استفاده از روش گرادیان، ضرایب خروجی تعیین می‌شوند. در لایه‌های بعدی، پس از مشارکت دادن خروجی‌های لایه‌های قبل، به عنوان بعدی جدید در کنار سایر ابعاد فضای ورودی، مجدداً خوشه‌بندی جهت تعیین ضرایب بخش مقدم قواعد و روش گرادیان برای تعیین رابطه‌ی خطی میان تالی قاعده با خروجی‌های لایه‌های قبل اعمال می‌شود. شمایی کلی از ساختار سیستم پیشنهادی در شکل (۲) نشان داده شده است.

۲-۲-۲- نحوه‌ی عمل کرد خوشه‌بندی مقاوم

اگر داده‌ها را در فضای اقلیدسی با بردار $\mathbf{X} = \{x_1, x_2, x_3, \dots, x_N\}^T, x_k \in \mathbb{R}^d$ که دارای d بعد باشند و بردار \mathbf{V} به صورت فاصله‌ی اقلیدسی میان نقاط x_k و مراکز خوشه‌ها C_i باشد که در آن کلاستر وجود دارد، $V = \|x_k - C_i\|^2 = \sqrt{\sum_{j=1}^d (x_{kj} - C_{ij})^2}$ تابع هدف فازی c -means با J تعریف می‌شود و در آن m اندیس فازی است.

$$\min J(U, V) = \sum_{i=1}^c \sum_{k=1}^n \mu_{ik}^m V(x_k, C_i) \quad (1)$$

که برای هر k داریم: $\sum_{i=1}^c \mu_{ik} = 1$,

$$v_i = \sum_{i=1}^n \mu_{ik} x_i / \sum_{i=1}^n \mu_{ik} \quad (2)$$

$$\mu_{ik} = \left[\sum_{j=1}^c (v_{ik} / v_{jk})^{1/(m-1)} \right]^{-1} \text{ for } \forall i, k \quad (3)$$

الگوریتم FCM به صورت تکراری اجرا می‌شود و باید حائز شرایط اولیه‌ای باشد که J را حداقل سازد. در هر اجرا، مرکز خوشه‌ها و عضویت‌های هر داده به خوشه‌ها به‌روز می‌شود. به منظور بهبود عمل کرد خوشه‌بندی FCM، در این مقاله از تابع هزینه‌ای که در [۲۶] ارائه گشته استفاده شده که به صورت زیر است:

$$\begin{aligned} \min J(U, \alpha, V) &= \sum_{i=1}^c \sum_{k=1}^n \mu_{ik} V(x_k, C_i) \\ &- r_1 \sum_{i=1}^c \sum_{k=1}^n \mu_{ik} \ln \alpha_k \\ &+ r_2 \sum_{i=1}^c \sum_{k=1}^n \mu_{ik} \ln \mu_{ik} \\ &- r_3 n \sum_{k=1}^n \alpha_k \ln \alpha_k \end{aligned} \quad (4)$$

۳-۲-۳- آموزش مدل

بخش مقدم قواعد با استفاده از خوشه‌بندی پیشنهاد شده آموزش می‌یابد. در بخش تالی، در هر لایه از یک TSK استفاده می‌شود. در هر لایه، خروجی لایه‌های قبل به عنوان یک بعد جدید در کنار فضای ورودی به ازای قواعد $k = 1, \dots, K$:

Rule k: If X is A^k and y_1 is $A_{y_1}^k$ and ... y_{l-1} is $A_{y_{l-1}}^k$
then $y_{dp}^k = w_1^k y_1 + \dots + w_{dp-1}^k y_{dp-1} + b_1^k$

همان‌گونه که مشاهده می‌شود، خروجی لایه‌ی قبل در هر دو بخش مقدم و تالی حضور دارد و از این منظر می‌توان چنین ساختاری را سیستم فازی مبتنی بر قواعد TSK با تالی غیرخطی در نظر گرفت. با تکرار چند لایه از این ساختار، سیستم نهایی، یک سیستم سلسله‌مراتبی یا عمیق تلقی می‌شود. که لایه‌های قبل در هر دو بخش لایه‌های بعد مشارکت دارند. برای آموزش تالی قواعد هر لایه، با استفاده از تابع خطای زیر و روش گرادیان کاهشی می‌توان به ضرایب مناسب دست یافت.

$$E_{dp} = - \sum_{i=1}^N y_{actual}^i \log(\sigma(W_{dp} Y_g^i)) \quad (13)$$

که در این رابطه، N تعداد مشاهدات، σ تابع ماکزیم-نرم به صورت $\sigma(z) = e^z / (1 + \sum e^z)$ و ماتریس ضرایب مجهول می‌باشد و به صورت زیر در نظر گرفته می‌شود.

$$W_{dp} = [(w_1^1, \dots, w_{dp-1}^1, b_1^1), \dots, (w_1^K, \dots, w_{dp-1}^K, b_1^K)] \quad (14)$$

و نیز

$$Y_g^i = [\tilde{y}_{dp}^1(i), \dots, \tilde{y}_{dp}^K(i)] \\ \tilde{y}_{dp}^k(i) = \mu_l^k(X_i) * \mu_l^k(y_i) * [y_1(i), \dots, y_{dp-1}(i), 1] \quad (15)$$

که بعد بردار واحد 1 توسط بعد خروجی تعیین می‌شود. مقادیر بهینه برای بردار وزن‌ها به روش گرادیان نزولی به صورت زیر محاسبه می‌شوند:

$$W_{dp}(t+1) = W_{dp}(t) - \eta \frac{\partial E_{dp}}{\partial W_{dp}} \\ = W_{dp}(t) + \eta \left(\sum_{i=1}^N \frac{\partial \sigma(W_{dp} Y_g^i)}{\partial W_{dp}} (y_{actual}^i - \sigma(W_{dp} Y_g^i)) \right) \\ = W_{dp}(t) + \eta \left(\sum_{i=1}^N Y_g^i (y_{actual}^i - \sigma(W_{dp} Y_g^i)) \right) \quad (16)$$

که در این رابطه، η ضریب یادگیری و مقداری ثابت است.

$$\mu_{ik} = \exp\left(\frac{-d_{ik}^2 + r_1 \ln \alpha_k}{r_2}\right) / \sum_{t=1}^c \exp\left(\frac{-d_{ik}^2 + r_1 \ln \alpha_k}{r_2}\right) \quad (5)$$

۲- با استفاده از معادلات زیر مقادیر $r_1^{(t)}$ و $r_2^{(t)}$ به روز رسانی شود.

$$r_1^{(t)} = e^{-t/10} \quad (6)$$

$$r_2^{(t)} = e^{-t/100} \quad (7)$$

۳- با استفاده از مقادیر $\mu_{ik}^{(t)}$ و $a_k^{(t-1)}$ معادله‌ی زیر مقدار $a_k^{(t)}$ به روز رسانی شود.

$$\alpha_k^{(new)} = \frac{1}{n} \sum_{i=1}^n \mu_{ik} + \frac{r_3}{r_1} \alpha_k^{(old)} (\ln \alpha_k^{(old)} - \sum_{t=1}^c \alpha_t^{(old)} \ln \alpha_t^{(old)}) \quad (8)$$

۴- با استفاده از مقادیر $a_k^{(t-1)}$ و $a_k^{(t)}$ معادله‌ی زیر، مقدار $r_3^{(t)}$ به روز رسانی شود.

$$r_3 = \min \left(\frac{\sum_{k=1}^c \exp(-\eta n |\alpha_k^{(new)} - \alpha_k^{(old)}|)}{1 - \max_{1 \leq k \leq n} \frac{1}{n} \sum_{k=1}^n \mu_{ik}} - \max_{1 \leq k \leq n} \alpha_k^{(old)} \sum_{t=1}^c \alpha_t^{(old)} \ln \alpha_t^{(old)} \right) \quad (9)$$

۵- خوشه‌هایی که $a_k^{(t)} \leq 1/n$ هستند حذف شده و مقدار $c^{(t)}$ به روز شود.

$$c^{(new)} = c^{(old)} - \left| \left\{ \alpha_k^{(new)} \mid \alpha_k^{(new)} < \frac{1}{n} \right\} \right|, k = 1, 2, \dots, c^{(old)} \quad (10)$$

مقادیر $a_k^{(t)}$ و $\mu_{ik}^{(t)}$ با استفاده از معادلات زیر به روز شوند.

$$a_k^* = a_k^* / \sum_{t=1}^c \alpha_t^{(new)} \quad (11)$$

$$\mu_{ik}^* = \mu_{ik}^* / \sum_{t=1}^c \mu_{it}^{(new)} \quad (12)$$

اگر $t \geq 100$ و $c^{(t-100)} - c^{(t)} = 0$ باشد، مقدار $r_3^{(t)} = 0$ گردد.

۶- با استفاده از مقادیر $\mu_{ik}^{(t)}$ و $c^{(t)}$ معادله‌ی (۳) مقدار $v_k^{(t)}$ به روز رسانی شود.

شرط توقف: دو مقدار متوالی $v_k^{(t)}$ و $v_k^{(t-1)}$ مقایسه گردند. اگر $|v_k^{(t)} - v_k^{(t-1)}| < \epsilon$ متوقف شود. در غیر این صورت اگر $t = t + 1$ و به ابتدای گام ۱ برود.

۴-۲- خروجی هر لایه Z_{dp} با استفاده از رابطه‌ی زیر ساخته

$$\text{شود: } Z_{dp} = \sigma(W_{dp} Y_g)$$

۵-۲- ورودی لایه‌ی بعدی از طریق روابط زیر محاسبه شود:

$$X_{dp} = [X Z_{dp}]$$

$$dp = dp + 1$$

۳- مقادیر آموزش یافته‌ی مراکز خوشه‌ها و ضرایب خروجی قواعد در هر لایه بازگردانده شود.

۴-۲-۳- ترکیب نتایج

بعد از آموزش لایه‌های مختلف مدل، از ترکیب نتایج لایه‌های مختلف به عنوان خروجی نهایی استفاده می‌شود. برای ترکیب نتایج یا به عبارت دیگر جوشش^۱، روش‌های متنوعی وجود دارد. منظور از ترکیب در واقع فرایند ترکیب نتایج منابع مختلف در تصمیم‌گیری نهایی است. مطالعات بسیاری نشان می‌دهند که نتایج حاصل از ترکیب اطلاعات، بهتر از نتایج اطلاعات مجزا است. یکی از روش‌های جوشش، نظریه‌ی شواهد^۲ است [۲۷] که در این پژوهش از این روش برای ترکیب نتایج لایه‌های مختلف طبقه‌بند با تعبیر منابع مختلف بهره گرفته شده است. نظریه‌ی شواهد یک روش استنتاج مبتنی بر باور است که برای استنتاج از منابع مختلف مناسب می‌باشد. در این روش، مجموع باورها برای گزاره‌های مختلف نباید برابر مقدار واحد باشد، یعنی اجباری وجود ندارد که اگر برای گزاره‌ای باوری اختصاص نیافت، حتماً به نقیض آن اختصاص یابد. در این نظریه، تابع باور^۳ و تابع معقولیت^۴ به عنوان کرانه‌های بالا و پایین بازه‌ی باور در نظر گرفته می‌شود تا با استفاده از آن، عدم دقت و عدم قطعیت فرضیه مشخص گردد. این کار سبب تفکیک جهل در مورد یک فرضیه از عدم قطعیت آن می‌شود. در این نظریه، از انتساب باور پایه برای توزیع اطلاعات اولیه استفاده می‌شود که مشابه حالت احتمالاتی، به هر یک از گزاره‌های موجود عددی بین صفر و یک اختصاص می‌دهد.

$[0,1] \rightarrow 2^X$ که مقدار m برای گزاره‌ی A معادل با قسمتی از باور کلی است که به این گزاره اختصاص یافته و از طریق رابطه‌ی میان تابع باور و تابع معقولیت محاسبه می‌شود. قواعد کلی در خصوص تابع باور پایه به این صورت است که اول) هر گزاره احتمال وقوع بزرگ‌تر مساوی صفر داشته باشد $(m(A) \geq 0, A \in 2^X)$.

دوم) تمامی متغیرهای مساله در چارچوب دید وجود داشته باشند $(m(\emptyset) = 0)$.

نکته‌ی قابل توجه در بخش آموزش مقدم آن است که در هر لایه، بعد فضای ورودی تغییر می‌یابد و با توجه به ابعاد خروجی لایه‌ی قبل که توسط تعداد کلاس‌ها تعیین می‌شود، بعد لایه‌ی بعد تعیین می‌گردد. در روش خوشه‌بندی مورد استفاده، تعداد قواعد توسط تعداد خوشه‌ها تعیین می‌شود و برای به دست آوردن درجه‌ی عضویت هر قاعده نیاز است که درجه‌ی عضویت فضای ورودی برای هر قاعده از درجه‌ی عضویت خروجی‌های لایه‌های قبل تفکیک شود. یعنی نیاز است که $\mu_i^k(X_i)$ و $\mu_i^k(Y_i)$ از یک‌دیگر تفکیک شوند. برای این منظور، پس از طی فرایند خوشه‌بندی، فضای خوشه‌های به دست آمده در دو زیرفضای ورودی و خروجی تصویر می‌شوند. فاصله‌ی دادگان در فضای ورودی به تصویر مراکز خوشه‌های به دست آمده در زیرفضای مربوطه $\mu_i^k(X_i)$ را می‌سازد و فاصله‌ی دادگان در فضای خروجی لایه‌ی قبل تا تصویر مراکز خوشه‌ها در زیرفضای خروجی، $\mu_i^k(Y_i)$ را می‌سازد.

الگوریتم ۲: نحوه‌ی آموزش شبکه‌ی فازی عمیق

ورودی: داده‌ی آموزش

$$X_{train} = \{x_1, x_2, x_3, \dots, x_N\}^T$$

مجموعه‌ی خروجی متناظر

$$Y_{actual} = \{y_{actual}^1, \dots, y_{actual}^N\}^T$$

که $y_i \in \mathbb{R}^m$ و $x_i \in \mathbb{R}^d$

خروجی: تابع پیش‌بینی مدل و قواعد فازی در هر لایه (شامل مراکز خوشه‌ها و ضرایب خروجی قواعد در هر لایه)

۱- مقادیر اولیه به صورت $dp=1$ و $X_{dp} = X$ قرار گیرد.

۲- تا زمانی که $dp = DP$ گردد مراحل زیر تکرار شود:

۱-۲- برای نمونه‌های هر کلاس، خوشه‌بندی بخش مقدم با استفاده از الگوریتم ۱ مراکز خوشه‌ها و تعداد آن‌ها تعیین شود:

۲-۲- ماتریس خروجی قواعد $\mu_i^k(X_i)$ از طریق رابطه‌ی زیر محاسبه شود:

$$\mu_i^k(X_i) = \left[\sum_{j=1}^c (\vartheta_{ik} / \vartheta_{jk}) \right]^{-1} \text{ for } \forall i, k$$

که در این رابطه ϑ_{ik} فاصله‌ی نمونه‌ی i -ام از مرکز خوشه‌ی قاعده‌ی k -ام می‌باشد.

۲-۳- بردار وزن خروجی W_{dp} در هر لایه با روش گرادیان کاهش‌ی به صورت تکرار شونده از رابطه‌ی (۱۶) به دست آید.

^۱ Belief Function

^۲ Plausibility Function

^۳ Fusion

^۴ Evidence Theory

۲-۲- ماتریس خروجی Z_{dp} با استفاده از رابطه‌ی زیر محاسبه شود: $Z_{dp} = \sigma(W_{dp}Y_g)$ که در این رابطه Y_g از رابطه‌ی (۱۵) به دست می‌آید.

۲-۳- داده‌های لایه‌ی بعدی به صورت زیر ساخته شود:

$$X_{dp} = [X \ Z_{dp}]$$

$$dp = dp + 1 \quad ۲-۴$$

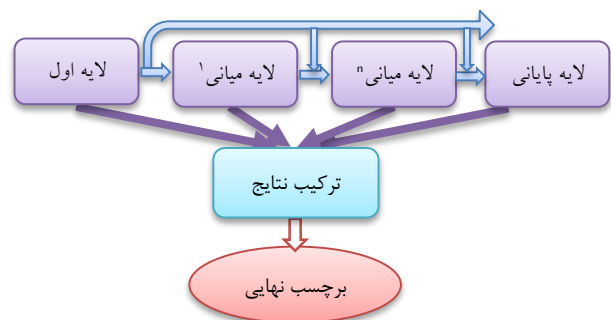
۳- مقدار خروجی هر لایه Z_{dp} بازگردانده شود و مقدار خروجی نهایی با استفاده از اطلاعات خروجی تمام لایه‌ها و قاعده‌ی ترکیب دمپستر-شافر مطابق رابطه‌ی (۱۷) محاسبه شود.

سوم) مجموع کل باورها برابر یک باشد $(\sum_{A \in 2^X} m(A) = 1)$. در میان انواع روش‌های ترکیب معرفی شده در تئوری شواهد، قاعده‌ی دمپستر^۱ در مواردی کاربرد دارد که شباهت زیادی بین باور حاصل از چند منبع متفاوت وجود دارد [۲۸].

$$m(A) = \frac{\sum_{C_i \cap \dots \cap C_p = A} \prod m_i(C_i)}{1 - \sum_{C_i \cap \dots \cap C_p = \emptyset} \prod m_i(C_i)} \quad (۱۷)$$

$A \neq \emptyset$

در این رابطه باورهای پایه‌ی $m_i(C_i)$ مربوط به منابع مختلف هستند و مخرج کسر بیان‌گر تناقض^۲ میان منابع مختلف است. ساختار تصمیم‌گیری مدل ارائه شده در شکل (۳) نشان داده شده است.



شکل (۳) - ساختار تصمیم‌گیری نهایی

۴- یافته‌ها و بحث

برای بررسی عمیق و دقیق عمل‌کرد مدل پیشنهادی، ابتدا به بررسی عمل‌کرد خوشه‌بند مقاوم استفاده شده در ساختار مدل پرداخته می‌شود. برای این مقایسه، به جای استفاده از RLFCM در شبکه، از FCM استفاده شد. در ساختار خوشه‌بند FCM نیاز است که تعداد خوشه‌ها در هر لایه از پیش تعیین شود. بنابراین برای چندین حالت مختلف از تعداد خوشه‌ها در FCM، مدل اجرا شد. در شکل (۴) نتایج مربوط به طبقه‌بندی چهار کلاس از سطوح مختلف افسردگی با استفاده از دادگان مورد استفاده برای ترکیب‌های مختلف خوشه‌بندی FCM (ستون‌های دوم تا هفتم از راست) در مقایسه با مدل اصلی یعنی استفاده از RL-FCM در شبکه‌ی اصلی (اولین ستون از راست) ارائه شده است. در روش خوشه‌بندی RL-FCM، تعداد خوشه‌ها در هر لایه توسط خود الگوریتم تعیین می‌شود که به صورت متوسط روی ۳۰ تکرار (۱۸/۲۳) گزارش شده است، اما برای روش FCM به ازای تعداد متفاوت خوشه‌ها، نتایج نشان داده شده است (اعداد داخل کروشه به ترتیب بیان‌گر تعداد خوشه‌ها در لایه‌ی اول تا سوم است).

نتایج شکل (۴) چندین نکته را بیان می‌دارد:

- ۱- مقایسه‌ی نتایج به دست آمده بیان‌گر قدرت بالای خوشه‌بند RL-FCM در مدل فازی عمیق در مقایسه با حالتی است که از خوشه‌بند FCM در مدل فازی عمیق استفاده شود و صحت بالا در دادگان تست بیان‌گر این قابلیت است.
- ۲- نتایج در تعیین خودکار تعداد خوشه‌ها (استفاده از الگوریتم RL-FCM) مناسب‌تر از نتایج حاصل از پیکره‌بندی‌های مختلف FCM می‌باشد. اما نکته‌ای که وجود دارد این است که ما در

۳-۲-۵- ارزیابی مدل

بعد از آموزش مدل با الگوریتم ۲، مدل قادر به پیش‌بینی نمونه‌های ناشناخته‌ی تست خواهد بود. روند پیش‌بینی در الگوریتم ۳ نشان داده شده است. الگوریتم ۳ روند پیش‌بینی بر حسب ساختار آموزش‌یافته و قواعد استخراج‌شده را به صورت خلاصه بیان می‌نماید.

الگوریتم ۳: روند پیش‌بینی

ورودی: داده‌ی تست $X_{test} = \{x_1, x_2, x_3, \dots, x_N\}^T$

مراکز خوشه‌ها v_i و ضرایب وزنی در هر لایه W_{dp}

خروجی: مجموعه‌ی خروجی متناظر

$$y_{actual} = \{y_{actual}^1, \dots, y_{actual}^N\}^T$$

۱- مقادیر اولیه به صورت $dp=1$ و $X_{dp} = X_{test}$ قرار گیرد.

۲- تا زمانی که $dp = DP$ گردد مراحل زیر تکرار شود:

- ۱-۲- با استفاده از خوشه‌های به دست آمده در مرحله‌ی آموزش برای هر لایه، ماتریس عضویت μ_{dp} برای تعلق دادگان تست به هر کدام از خوشه‌های قواعد محاسبه شود

^۲ Contradiction

^۱ Dempster's rule of combination

سمت راست) با الگوریتم RLFCM استفاده شده در ساختار فازی عمیق پیشنهادی (اولین ستون از سمت راست) به منظور نمایش ضرورت ترکیب نتایج لایه‌های مختلف شبکه، نتایج هر لایه به صورت تفکیکی بررسی شده است. در جدول (۲) نتایج حاصل از روش‌های مختلف ترکیب خروجی لایه‌های سیستم عمیق پیشنهادی نشان داده شده است.

جدول (۲) - مقایسه‌ی نتایج حاصل از لایه‌های مختلف سیستم عمیق با روش ترکیب اطلاعات در لایه‌ها در مدل پیشنهادی

خروجی لایه‌های سیستم عمیق پیشنهادی	صحت در فاز آموزش	صحت در فاز تست
تک لایه ابتدایی	۵۱.۰۵٪ (۹.۱۶±٪)	۴۵.۱۲٪ (۱۰.۴۹±٪)
تک لایه میانی	۵۲.۵۹٪ (۶.۷۰±٪)	۴۸.۴۷٪ (۶.۰۷±٪)
تک لایه انتهایی	۴۵.۱۶٪ (۹.۳۱±٪)	۴۵.۷۷٪ (۷.۲۳±٪)
ترکیب خروجی لایه‌های سیستم	۵۲.۴۴٪ (۶.۰۳±٪)	۴۹.۰۱٪ (۵.۵۲±٪)

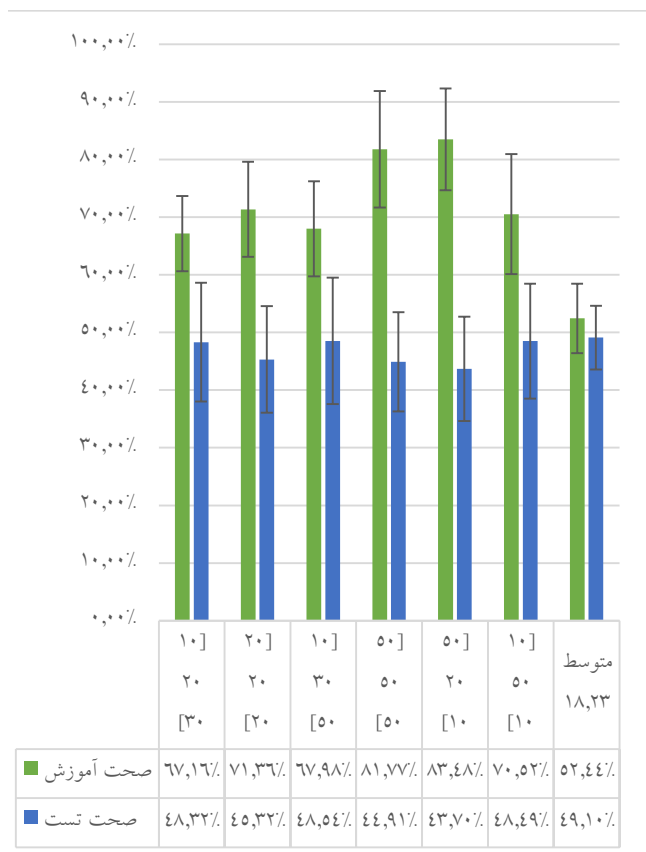
نتایج حاصل از جدول (۲) نشان می‌دهد که با وجود این که ساختار سیستم عمیق پیشنهادی به نوعی روشی برای ترکیب نتایج به دست آمده در لایه‌های مختلف است، بهره‌گیری از ترکیب نتایج لایه‌های مختلف در مقایسه با نتایج تک‌لایه، گزینه‌ی مناسب‌تری است.

در نمودار شکل (۵) به مقایسه‌ی عمل کرد کلی روش پیشنهادی با چند روش مطرح در طبقه‌بندی، نظیر شبکه‌ی عصبی جلوسو (با تعداد لایه‌ی پنهان ۲۰)، درخت تصمیم، بیز ساده، ماشین بردار پشتیبان و آنالیز افتراقی خطی پرداخته شده است. برای دو طبقه‌بند شبکه‌ی عصبی و درخت تصمیم، مشکل بیش‌برازش^۱ رخ داد که برای رفع این مشکل، دادگان آموزش را به حدی کاهش دادیم که با افزایش بیش‌تر تعداد داده‌های تعلیم، صحت آموزش افزایش اما صحت تست کاهش می‌یافت. نتایج حاصل از این مقایسه بیان‌گر قدرت قابل توجه روش پیشنهادی نسبت به روش‌های متداول، هم از جهت صحت و هم از جهت پایداری می‌باشد. با وجود این که طبقه‌بندی نظیر درخت تصمیم یا شبکه‌ی عصبی در فاز آموزش نتایج بالاتری را کسب کرده‌اند، اما با مقایسه‌ی نتایج در فاز تست، مشخص می‌گردد که طبقه‌بندها دچار مشکل بیش‌برازش شده‌اند که

این‌جا تمام حالت‌های FCM را بررسی نکردیم و ممکن است با پیکره‌بندی‌های دیگری از FCM (که بررسی نشده است) به نتایج بهتری نسبت به روش مورد استفاده دست یابیم، اما نقطه‌ی قوت این روش در تعیین خودکار تعداد خوشه‌ها برای جواب مناسب است که ما را از اجرای مکرر و طراحی‌های متفاوت خوشه‌بندی به ازای تعداد مختلف خوشه‌ها بی‌نیاز می‌سازد.

۳- با مقایسه‌ی تعداد قواعد به دست آمده در روش مورد استفاده (که به صورت متوسط ۱۸ قاعده‌ی اجراهای مختلف است) ضرورت به کارگیری تعداد مناسب قواعد برای آموزش مدل دریافت می‌شود.

۴- خطوط عمودی مشکی رنگ روی ستون‌های شکل، بیان‌گر انحراف معیار نتایج به ازای تکرارهای مختلف است. انحراف معیار صحت در دادگان تست در حالتی که از RL-FCM در مدل فازی عمیق استفاده شده در مقایسه با مدل ترکیبی با FCM کم‌تر و نشان‌گر پایداری بیش‌تر است (کم‌ترین مقدار انحراف معیار یعنی ۵/۵٪).



شکل (۴) - مقایسه‌ی نتایج پیکره‌بندی‌های مختلف خوشه‌بندی در ساختار فازی عمیق پیشنهادی (ستون دوم تا هفتم از

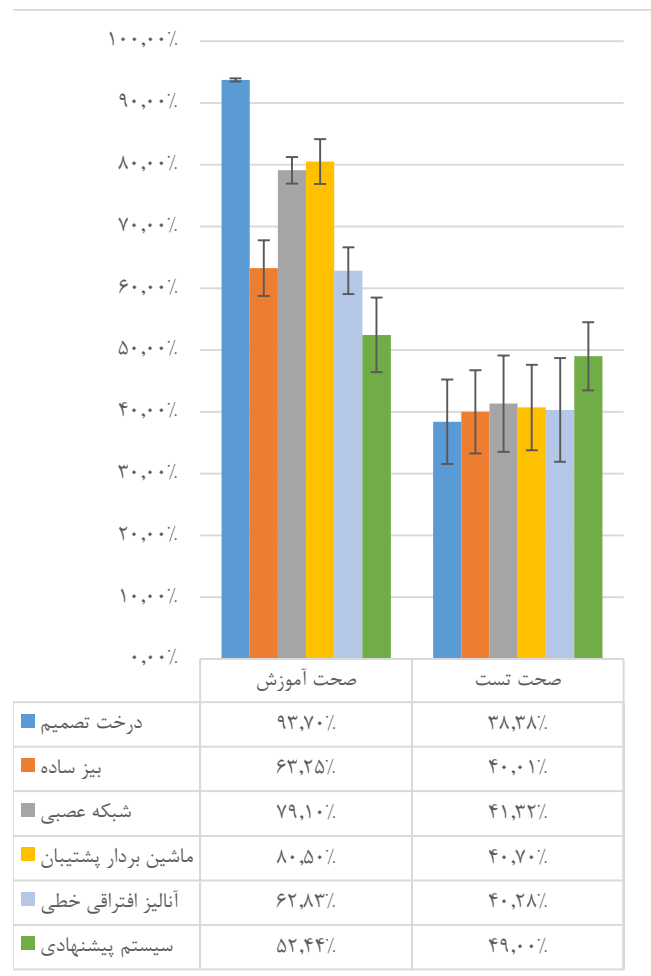
^۱ Overfitting

برای مشاهده‌ی بهتر توان‌مندی‌ها و عمل‌کرد مدل فازی عمیق پیشنهادی، به جداسازی افراد سالم (نمره‌ی زیر ۱۰ در پرسش‌نامه) و افراد دارای افسردگی (نمره‌ی بالای ۱۱) پرداخته و نتایج به دست آمده با رویکرد ارائه شده در [۸] مقایسه می‌شود. جداسازی افراد سالم از افسرده در [۸] با استفاده از طبقه‌بند خطی و مجموعه‌ی ویژگی‌های غیرخطی نظیر نمای لیاپانوف، بعد همبستگی، تحلیل نوسانات کاهش‌یافته و انتخاب ویژگی بهینه با الگوریتم ژنتیک به روش ارزیابی تقاطعی یکی-بیرون^۱ صورت گرفت و به نتیجه‌ی صحت جداسازی ۹۰٪ رسید. در این پژوهش نیز با استفاده از ویژگی‌های اشاره شده در بخش ۳-۱ و سیستم پیشنهادی بدون استفاده از بهینه‌سازی ژنتیک به صحت ۹۱/۹۲٪ دست یافتیم که برتری روش پیشنهادی را نشان می‌دهد. در پایان، لازم به ذکر است که در مدل فازی عمیق پیشنهادی هیچ‌گونه پارامتری بهینه نگردیده و با بهینه‌سازی پارامترها به پاسخ مناسب‌تر می‌توان دست یافت.

۵- نتیجه‌گیری

در این مقاله، یک مدل فازی عمیق با ساختار چند لایه و قواعد تفسیرپذیر فازی ارائه شد که در کاربرد تشخیص سطوح افسردگی (ضعیف، ملایم، متوسط و شدید) با استفاده از سیگنال EEG به کار گرفته شد. قواعد فازی استخراج شده در ساختار پیشنهادی در این مقاله، با استفاده از الگوریتم خوشه‌بندی مقاومی استخراج می‌شوند که قادر است بدون نیاز به دانش اولیه درباره‌ی تعداد خوشه‌ها، تعداد بهینه‌ای خوشه استخراج کند. ساختار لایه‌های بالاتر این سیستم عمیق به صورت ترکیبی از دانش خروجی لایه‌ی قبل به عنوان یک بعد اضافه شده به ورودی در نظر گرفته می‌شود و قواعد استخراج شده در لایه‌های بالاتر حاصل ترکیب دانش آموزش یافته در لایه‌های پیشین در کنار فضای ورودی است. حضور دانش لایه‌های پیشین در دو بخش از لایه‌های بعدی، بخش مقدم و بخش تالی، قابل توجه است. افزون بر موارد ذکر شده، در تصمیم‌گیری نهایی، دانش در لایه‌های مختلف ترکیب شده و تصمیم‌نهایی اتخاذ می‌شود. نتایج به دست آمده از مقایسه‌ی کارایی این مدل در مقایسه با روش‌های دیگر طبقه‌بندی نشانه‌ی برتری عمل‌کرد آن در دادگان مورد استفاده است.

قابل قبول نیست. در جدول (۳) نمونه‌ای از قواعد به دست آمده توسط مدل فازی عمیق در هر لایه بیان شده است. همان‌گونه که مشاهده می‌شود، در لایه‌ی اول، مقدم هر قاعده به صورت تعلق فضای ورودی و تالی قاعده به صورت TSK مرتبه‌ی صفر یعنی یک ضرب، بیان شده است. در لایه‌ی دوم، مقدم قواعد دو بخش دارند: مرکز خوشه متعلق به فضای ورودی و مرکز خوشه متعلق به فضای خروجی لایه‌ی اول، که به صورت جداگانه در شکل بیان شده است. در تالی قواعد لایه‌ی دوم، خروجی‌های هر کلاس به صورت یک رابطه‌ی خطی با خروجی لایه‌ی اول کلاس بیان شده که به نوعی TSK مرتبه‌ی اول محسوب می‌شود و در نهایت قاعده‌ی لایه‌ی سوم که در مقدم شامل سه خوشه متعلق به فضای ورودی، فضای خروجی لایه‌ی اول و فضای خروجی لایه‌ی دوم است و در خروجی به صورت رابطه‌ی خطی برحسب خروجی‌های لایه‌های اول و دوم بیان می‌شود.



شکل (۵) - مقایسه‌ی نتایج طبقه‌بندی‌های مختلف با ساختار فازی عمیق پیشنهادی

^۱ Leave one out

جدول (۳) - نمونه‌ای از قواعد به دست آمده در ساختار فازی عمیق پیشنهادی

نمونه ای از قاعده در لایه اول				
مقدم قاعده				مقدم قاعده
اگر فضای ورودی متعلق به خوشه زیر باشد: $C = [0.3778; 0.6608; -0.3105; 0.4207; 0.0206; -0.4303; -0.1508; 0.1110; -0.1112; 0.4967]$				
خروجی کلاس افسردگی شدید	خروجی کلاس افسردگی متوسط	خروجی کلاس افسردگی ملایم	خروجی کلاس افسردگی ضعیف	تالی قاعده
$y1_4 = -0.6780$	$y1_3 = -4.8759$	$y1_2 = 1.1253$	$y1_1 = -1.7265$	
نمونه ای از قاعده در لایه دوم				
مقدم قاعده				مقدم قاعده
اگر فضای ورودی متعلق به خوشه زیر باشد: $C = [0.3061; -0.0751; 0.0446; -0.3733; -0.9502; -0.05812; 0.01150; 0.2188; -0.6419; -0.1630]$ و خروجی‌های لایه اول متعلق به خوشه زیر باشد: $C' = [0.2599; 0.0952; 0.1273; 0.2700]$				
خروجی کلاس افسردگی شدید	خروجی کلاس افسردگی متوسط	خروجی کلاس افسردگی ملایم	خروجی کلاس افسردگی ضعیف	تالی قاعده
$y2_4 = +0.6034 \times y1_4$ 3.6637	$y2_3 = -0.4375 \times y1_3$ -3.9336	$y2_2 = -0.1991 \times y1_2$ -4.0014	$y2_1 = -0.0428 \times y1_1$ -3.7559	
نمونه ای از قاعده در لایه سوم				
مقدم قاعده				مقدم قاعده
اگر فضای ورودی متعلق به خوشه زیر باشد: $C = [0.0914; 0.8012; 0.1593; 0.2958; -0.1163; -0.3869; -0.4580; 0.0120; 0.0402; 0.1946]$ و خروجی‌های لایه اول متعلق به خوشه زیر باشد: $C' = [0.1419; 0.1328; 0.2407; 0.2540]$ و خروجی‌های لایه دوم متعلق به خوشه زیر باشد: $C'' = [0.4837; 0.4436; 0.4309; 0.4724]$				
خروجی کلاس افسردگی شدید	خروجی کلاس افسردگی متوسط	خروجی کلاس افسردگی ملایم	خروجی کلاس افسردگی ضعیف	تالی قاعده
$y3_4 = 0.5317 \times y1_4$ $+1.3753 \times y2_4$ +0.8232	$y3_3 = -1.5713 \times y1_3$ $-0.1875 \times y2_3$ -0.5334	$y3_2 = 0.3423 \times y1_2$ $-0.3398 \times y2_2$ -0.5135	$y3_1 = -0.8764 \times y1_1$ $-0.2339 \times y2_1$ +1.5398	

in depression for their diagnostic value," *BMC Med. Inform. Decis. Mak.*, vol. 15, no. 1, p. 108, Dec. 2015.

- [5] İ. Güler and E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," *J. Neurosci. Methods*, vol. 148, no. 2, pp. 113–121, Oct. 2005.
- [6] V. A. Grin-Yatsenko, I. Baas, V. A. Ponomarev, and J. D. Kropotov, "Independent component approach to the analysis of EEG recordings at early stages of depressive disorders," *Clin. Neurophysiol.*, vol. 121, no. 3, pp. 281–289, Mar. 2010.
- [7] S.-C. Liao, C.-T. Wu, H.-C. Huang, W.-T. Cheng, and Y.-H. Liu, "Major Depression Detection from EEG Signals Using Kernel Eigen-Filter-Bank Common Spatial Patterns," *Sensors*, vol. 17, no. 6, p. 1385, Jun. 2017.
- [8] B. Hosseinifard, M. H. Moradi, and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal," *Comput.*

۶- سپاس‌گزاری

نویسندگان از خانم بهشاد حسینی‌فرد برای ثبت سیگنال‌های مغزی و آقای یوسف محمدی برای پیش‌پردازش و استخراج ویژگی‌ها تشکر می‌کنند.

۷- مراجع

- [1] "Depression and Other Common Mental Disorders Global Health Estimates."
- [2] S. Tement, A. Pahor, and N. Jaušovec, "EEG alpha frequency correlates of burnout and depression: The role of gender," *Biol. Psychol.*, vol. 114, pp. 1–12, Feb. 2016.
- [3] W. Mumtaz, A. S. Malik, M. A. M. Yasin, and L. Xia, "Review on EEG and ERP predictive biomarkers for major depressive disorder," *Biomed. Signal Process. Control*, vol. 22, pp. 85–98, Sep. 2015.
- [4] M. Mohammadi *et al.*, "Data mining EEG signals

- [18] R. Davoodi and M. H. Moradi, "Mortality prediction in intensive care units (ICUs) using a deep rule-based fuzzy classifier," *J. Biomed. Inform.*, vol. 79, 2018.
- [19] A. Beck, R. Steer, G. B.-S. Antonio, and undefined 1996, "Beck depression inventory-II," *m.blog.naver.com*.
- [۲۰] ب. حسینی. فرد، "تشخیص و جداسازی بیماران افسرده از افراد سالم با استفاده از سیگنال الکتروانسفالگرام"، پایان نامه کارشناسی ارشد، دانشگاه صنعتی امیرکبیر، ۱۳۸۹.
- [21] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence Plots of Dynamical Systems," *Europhys. Lett.*, vol. 4, no. 9, pp. 973–977, Nov. 1987.
- [22] U. R. Acharya, S. V. Sree, S. Chattopadhyay, W. Yu, and P. C. A. Ang, "Application of Recurrence Quantification Analysis for the Automated Identification of Epileptic EEG Signals," *Int. J. Neural Syst.*, vol. 21, no. 03, pp. 199–211, Jun. 2011.
- [23] F. Bahari and A. Janghorbani, "EEG-based emotion recognition using Recurrence Plot analysis and K nearest neighbor classifier," in *2013 20th Iranian Conference on Biomedical Engineering (ICBME)*, 2013, pp. 228–233.
- [24] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Phys. Rep.*, vol. 438, no. 5–6, pp. 237–329, Jan. 2007.
- [25] L.J.P. van der Maaten. "Accelerating t-SNE using Tree-Based Algorithms," *J. of Machine Learning Research*, vol. 15, pp. 3221-3245, 2014.
- [26] M.-S. Yang and Y. Nataliani, "Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters," *Pattern Recognit.*, vol. 71, pp. 45–59, Nov. 2017.
- [27] M.-S. Yang and Y. Nataliani, "Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters," *Pattern Recognit.*, vol. 71, pp. 45–59, Nov. 2017.
- [28] Shafer, "A Mathematical Theory of Evidence," Books on Demand, 1976.
- [29] M. Beynon, B. Curry, and P. Morgan, "The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modelling," *Omega*, vol. 28, no. 1, pp. 37–50, Feb. 2000.
- Methods Programs Biomed.*, vol. 109, no. 3, pp. 339–345, Mar. 2013.
- [9] S. R. I. Gabran, S. Zhang, M. M. A. Salama, R. R. Mansour, and C. George, "Real-time automated neural-network sleep classifier using single channel EEG recording for detection of narcolepsy episodes," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 1136–1139.
- [10] Hanshu Cai, Xiaocong Sha, Xue Han, Shixin Wei, and Bin Hu, "Pervasive EEG diagnosis of depression using Deep Belief Network with three-electrodes EEG collector," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016, pp. 1239–1246.
- [11] J. Zhang, C. Tao, and P. Wang, "A Review of Soft Computing Based on Deep Learning," in *2016 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICII)*, 2016, pp. 136–144.
- [12] S. Zhou, Q. Chen, and X. Wang, "Fuzzy deep belief networks for semi-supervised sentiment classification," *Neurocomputing*, vol. 131, pp. 312–322, May 2014.
- [13] Y.-J. Zheng, W.-G. Sheng, X.-M. Sun, and S.-Y. Chen, "Airline Passenger Profiling Based on Fuzzy Deep Machine Learning," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 12, pp. 2911–2923, Dec. 2017.
- [14] Y. Deng, Z. Ren, Y. Kong, F. Bao, and Q. Dai, "A Hierarchical Fused Fuzzy Deep Neural Network for Data Classification," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 1006–1012, Aug. 2017.
- [15] R. Zhang, F. Shen, and J. Zhao, "A model with Fuzzy Granulation and Deep Belief Networks for exchange rate forecasting," in *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 366–373.
- [16] S. Park, S. J. Lee, E. Weiss, and Y. Motai, "Intra- and Inter-Fractional Variation Prediction of Lung Tumors Using Fuzzy Deep Learning," *IEEE J. Transl. Eng. Heal. Med.*, vol. 4, pp. 1–12, 2016.
- [17] T. Zhou, F.-L. Chung, and S. Wang, "Deep TSK Fuzzy Classifier With Stacked Generalization and Triplely Concise Interpretability Guarantee for Large Data," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 5, pp. 1207–1221, Oct. 2017.