# مروری بر نظریه تشدید انطباقی[۱] (ART)

هدی جلال کمالی*
استادیار گروه مهندسی کامپیوتر، مجتمع آموزش عالی زرند، کرمان، ایران
* نویسنده مسئول: Hodajalalkamali@uk.ac.ir

**چکیده**

اکثر الگوریتم‌های یادگیری یا نمی‌توانند موارد بسیار نادر را بیاموزند، یا در یک دنیای غیرایستا (یعنی دنیایی که در آن ممکن است تمام قوانین بصورت غیر قابل پیش بینی در طی زمان تغییر کنند) و هنگام یادگیری سریع یعنی هنگامیکه یک رخداد فقط با یکبار ارائه آموخته می‌شود، ناپایدار می‌شوند. نظریه تشدید انطباقی این مسایل را حل می کند. این نظریه که بر اساس پردازش اطلاعات شناختی بشر توسعه یافته است، شامل مدل های با سرپرست و بدون سرپرست می باشد. در این مقاله اصول یادگیری بر پایه انطباق (در مقابل یادگیری بر پایه خطا در پرسپترون) و یادگیری سریع پایدار در ART ، به علاوه نحوه کارکرد برخی از مدل های آن شامل ART۱،ARTMAP و dART ارائه می شود. کاربردهای گسترده این نظریه شامل تشخیص و طبقه بندی در حوزه های فنی، زیستی و پزشکی است که عناوین برخی از آنها در این مقاله آورده می شود.

**کلمات کلیدی**

نظریه تشدید انطباقی (ART)، شبکه عصبی، یادگیری پایدار، ART۱، ARMTMAP، dART

## ۱- مقدمه

نظریه تشدید انطباقی (ART) در سال ۱۹۷۶ توسط گراسبرگ وکارپنتر بر اساس آزمایشاتی در حوزه بینایی، شنوایی و یادگیری تقویتی به عنوان یک نظریه پردازش اطلاعات شناختی بشر معرفی شد (۱و۲). یادگیری بدون سرپرست[۲] و با سرپرست[۳] در این نظریه بصورت یک سری از مدل‌های شبکه عصبی بلادرنگ توسعه یافته است (۳و۴). مدل‌های یادگیری بدون سرپرست برای الگوهای ورودی دودویی، ART۱ (۵) و برای الگوهایی ورودی پیوسته ART فازی نامیده می شوند (۶). مدل‌های ARMTMAP (۷) دو ماژول بدون سرپرست را ترکیب نموده تا یادگیری با سرپرست را انجام دهند.

سیستم‌های ART و dART بخشی از خانواده رو به رشد مدلهای شبکه خودسازماندهی هستند که ویژگی‌های بازخوردِ توجهی و یادگیری پایدار را بطور برجسته نشان می‌دهند. تفاوتی که ART با مدلهای شبکه خودسازماندهی دارد این است که در آنها تنها یادگیری پایین به بالا انجام می‌شود. تحلیلی که روی مدل نقشه ویژگی خودسازماندهی[۴] (۱۰-۸)، یادگیری رقابتی انجام گرفت نشان داد که یادگیری تداعیگرانه و مهارجانبی که اساس طبقه بندی در این مدل هستند، در پاسخ به مجموعه اختیاری از داده‌های ورودی، موجب عدم پایداری[۵] می‌شود. اما ترکیب آن با تمرکزِ توجه، جستجوی حافظه و انطباق بالا به پائین (که در ART انجام می‌شود) موجب یادگیری پایدار در یک محیط با مجموعه ورودی‌های اختیاری می‌شود (۱و۲).

---
۱ Adaptive Resonance Theory
۲ unsupervised
۳ supervised
۴ Self-Organized Feature Map
۵ Instability

از زمان معرفی این نظریه انواع بسیاری از شبکه‌های بدون سرپرست و با سرپرست برای کاربردهای فنی و تحلیل‌های بیولوژیکی ارائه شده است. حوزه‌های کاربرد فنی شامل طراحی و ساخت صنعتی، کنترل ربات‌های متحرک، تشخیص چهره، حس کردن[۱] از راه دور، طبقه بندی پوشش زمین، تشخیص هدف، تشخیص پزشکی، تحلیل الکتروکاردیوگرام (ECG)، تصدیق امضاء، نظارت بر خرابی ابزار، تحلیل شیمیایی، طراحی مدار، تحلیل پروتئین/DNA ، تشخیص بصری شیء سه بعدی، تحلیل موسیقی، تشخیص زمین لرزه، سونار (دستگاه کاشف زیردریایی بوسیله امواج صوتی) و رادار است (۱۳-۱۱). کاربردهای پیاده سازی سیستم‌های ART برروی تراشه‌های VLSI توسط سرانو-گوتاردونا[۲] و همکاران (۱۴) مورد بحث قرار گرفته است. این کاربردها از قابلیت سیستم‌های ART در یادگیری طبقه بندی پایگاه داده‌های بزرگ بصورت پایدار، سنجیدن میزان اطمینان در یک طبقه بندی و تمرکزِ توجه بر طبقه بندی‌های برجسته که سیستم بر اساس تجربه‌اش آنها را مهم می‌پندارد بهره می‌گیرد. حافظه‌های ART همچنین به یک مجموعه ناپیدا از قوانین IF-THEN ترجمه می‌شوند که فرآیند تصمیم گیری را مشخص کرده و می‌توانند برای انتخاب ویژگی به کار روند. به این ترتیب ادعای برخی دانشمندان علوم شناختی مبنی بر اینکه مدلهای شبکه عصبی نمی‌توانند رفتارهای بر پایه قانون را بیاموزند اشتباه است همانند این ادعا که شبکه‌های عصبی نمی‌توانند سمبل‌ها را بیاموزند.

در ادامه این مقاله، یادگیری بر پایه انطباق که اساس یادگیری در ART است و مقایسه آن با یادگیری برپایه خطا که اساس یادگیری پس- انتشار[۳] خطاست و همچنین یادگیری سریع در بخش۲-۱ ، معمای پایداری- انعطاف پذیری در بخش۲-۲ و پارامتر مراقبت[۴] که یک پارامتر آزاد در ART است در بخش ۲-۳ آورده می‌شود. سه مدل ART یعنی ART۱، ARTMAP و dART به ترتیب در بخشهای۳، ۴ و ۵ به تفضیل توضیح داده می‌شوند و در نهایت بخش ۶ یک نتیجه گیری از مطالب بیان شده ارائه می نماید.

### ۲-۱ یادگیری بر پایه انطباق[۵] ، یادگیری بر پایه خطا و یادگیری سریع پایدار

یک ویژگی اساسی تمام سیستم‌های ART فرآیند انطباق الگو است که در آن یک ورودی خارجی با حافظه داخلی یک کد فعال مقایسه می‌شود. فرآیند انطباق ART یا به یک حالت تشدید می‌انجامد که به اندازه کافی به درازا می‌کشد تا یادگیری انجام شود یا به یک جستجوی موازی حافظه. اگر جستجوی حافظه به یک کد پایدار منجر شود بازنمایی حافظه یا بدون تغییر می‌ماند یا اطلاعات جدید را از بخش‌های منطبق با حافظه و ورودی جاری با آن ترکیب می‌کند. اگرجستجو به یک کد جدید بینجامد ، بازنمایی حافظه، ورودی جاری را یاد می‌گیرد. این فرایند یادگیری بر پایه انطباق، اساس پایداری کد ART است. یادگیری بر پایه انطباق تنها زمانی اجازه می‌دهد مقدار حافظه‌ها تغییر کند که ورودی از دنیای خارجی به اندازه کافی به انتظارات داخلی نزدیک باشد یا هنگامیکه چیز کاملاً جدیدی رخ دهد. این ویژگی موجب می‌شود سیستم‌های ART به خوبی برای مسائلی که مستلزم یادگیری بر خط پایگاه داده‌های بزرگ و به روز شونده هستند مناسب باشد.

یادگیری بر پایه انطباق مکمل یادگیری بر پایه خطاست، که به یک عدم انطباق به این صورت پاسخ می‌دهد که حافظه‌ها را تغییر داده به گونه ای که تفاوت بین خروجی هدف و خروجی واقعی کاهش یابد نه اینکه به دنبال یک انطباق بهتر بگردد. یادگیری بر پایه خطا بصورت طبیعی برای مسائلی مثل کنترل انطباقی و یادگیری نقشه‌های حسی- حرکتی مناسب است

---

۱ sensing
۲ Serrano-Gotarredona
۳ back-propagation
۴ vigilance
۵ Match-based Learning

که مستلزم انطباق مداوم با آمارهای موجود هستند. شبکه‌های عصبی که یادگیری بر پایه خطا را به کار می‌گیرند شامل پس- انتشار و دیگر پرسپترون‌های چند لایه[۱] (MLP) هستند (۴).

ART مشکل گیرکردن در مینیمم محلی را ندارد. گراسبرگ وکارپنتر بصورت ریاضی ثابت کردند که ورودی‌های آشنا بدون هیچ فرآیند جستجویی، بهترین انطباق کلی[۲] را با دسته متناظرشان پیدا می‌کنند (۵). در حالیکه ورودی‌های نا آشنا با راه اندازی فرآیند جستجوی حافظه که توسط سیستم سوگیری[۳] انجام می‌شود، بدنبال طبقه‌های بهتر می‌گردند و این فرآیند آنقدر ادامه می‌یابد تا ظرفیت حافظه تمام شود. در نتیجه ART مشکل مینیمم محلی که بسیاری از الگوریتم‌های دیگر نظیر پس- انتشار (۱۵و۱۶) دارند را حل می‌کند.

خاصیت کدگذاری با استفاده از "تمرکزِ توجه" بر روی ویژگی‌های بحرانی که برای پایداری کد اساسی است، یادگیری شبکه‌های ART را از MLPها متمایز می‌کند. MLPها معمولاً ورودی جاری را کدگذاری می‌کنند تا یک الگوی منطبق شده را بیاموزد و در نتیجه آزمایه‌های ورودی بسیاری را بکار می‌گیرد تا از فراموشی فاجعه بار اجتناب کند.

بسیاری از کاربردهای ART از یادگیری سریع استفاده می‌کنند که در آن وزن‌های انطباقی در پاسخ به هر الگوی ورودی به یک حالت تعادلی همگرا می‌شوند. یادگیری سریع باعث می‌شود که یک سیستم به سرعت با ورودی‌هایی منطبق شود که به ندرت رخ می‌دهند اما ممکن است مستلزم بازخوانی فوری و دقیق باشند. به یادآوردن جزئیات یک فیلم مهیج یک نمونه نوعی از یادگیری با یک آزمایه (با یک بار) است. یادگیری سریع حافظه‌هایی ایجاد می‌کند که به ترتیب ارائه ورودی بستگی دارند. بسیاری کاربردهای ART از این ویژگی برای بهبود دقت بهره برداری می‌کنند به این صورت که از میان چندین شبکه تعلیم داده شده رأی گیری کرده و رأی دهندگان در هر پیش بینی میزان اطمینان را تخمین می‌زنند.

### ۲-۲ معمای پایداری - انعطاف پذیری[۴]

یک مشکل ویژه در یادگیری شبکه‌ها، معمای پایداری-انعطاف پذیری است. این مساله بیان می‌کند که شبکه باید به اندازه کافی انعطاف پذیر باشد تا الگوهای ورودی جدید را ذخیره کند. در عین حال شبکه باید به اندازه کافی پایدار باشد تا از پاک شدن الگوهای رمزگذاری شده قبلی جلوگیری شود. گراسبرگ در سال ۱۹۸۰ این مساله را که مغز می‌تواند بسرعت و بصورت پایدار بیاموزد بدون اینکه آنچه که قبلا آموخته است را بطور فاجعه باری فراموش کند، معمای پایداری-انعطاف پذیری نامید (۱۷).

ART بصورت اتوماتیک یادگیری سریع، پایدار و افزایشی را هم در شرایط یادگیری با سرپرست و هم بدون سرپرست در یک دنیای پیچیده غیرایستا[۵] (یعنی دنیایی که در آن ممکن است تمام قوانین بصورت غیر قابل پیش بینی در طی زمان تغییر کنند) انجام می‌دهد. در مقابل بسیاری از الگوریتم‌های جاری یادگیری از یادگیری دسته ای استفاده می‌کنند که در آن تمام اطلاعاتی که باید در مورد دنیا آموخت یک باره در دسترس هستند، نظیر ماشین‌های بردار پشتیبانی[۶] (SVM) و پس-انتشار. اکثر الگوریتم‌های یادگیری در یک دنیای غیرایستا غیرپایدار می‌شوند یا نمی‌توانند موارد بسیار نادر را بیاموزند و یا هنگام یادگیری سریع یعنی هنگامیکه یک رخداد فقط با یکبار ارائه آموخته می‌شود، ناپایدار می‌شوند. از بعضی الگوریتم‌های یادگیری که دچار فراموشی فاجعه بار می‌شوند می توان به یادگیری رقابتی، نقشه‌های خودسازماندهی، Neocognitron و پس-انتشار اشاره نمود. ART بر این مشکلات غلبه می‌کند. راه حل ART برای حل مساله انعطاف

---

[۱] Mulit- Layer Perceptron
[۲] Globally best match
[۳] Orienting
[۴] Stability- Plasticity Dilemma
[۵] non-stationary
[۶] Support Vector Machine

پذیری-پایداری این است که انتظارات بالا به پائین و بازخوردی آموخته شده با الگوهای پائین به بالا مقایسه می‌شوند. در نتیجه توجه بر ترکیبی از ویژگی‌ها متمرکز می‌شود که در آن زمینه قابل پیش بینی هستند.

### ۲-۳ پارامتر مراقبت

معیار انطباق ART ، توسط پارامتر مراقبت تعیین می‌شود. این پارامتر حداقل کسری از ورودی را مشخص می‌کند که باید در الگوی منطبق شده وجود داشته باشد تا تشدید رخ دهد. اندازه کم پارامتر مراقبت تعمیم پذیری زیاد، طبقات بزرگ و حافظه‌های انتزاعی تر را موجب می‌شود، در حالیکه اندازه زیاد پارامتر مراقبت به تعمیم سازی اندک، طبقات ریزتر و حافظه با جزئیات بیشتر می‌انجامد.

در شبکه‌های بدون سرپرست ART ، مراقبت یک پارامتر آزاد است حال آنکه در شبکه‌های با سرپرست مراقبت متغیری است که بصورت داخلی کنترل می‌گردد که پس از صعود در پاسخ به یک خطای پیشگویانه یک جستجو را راه اندازی می‌نماید. از آنجا که میزان مراقبت سپس در میان آزمایه‌های یادگیری تغییر می‌کند، حافظه‌های یک سیستم منفرد ARTMAP معمولاً طیفی از درجات پالایش را نشان می‌دهند. با تغییر اندازه پارامتر مراقبت، یک سیستم می‌تواند هم طبقه‌های انتزاعی مثل صورت‌ها و سگ‌ها و هم نمونه‌های منفرد از این طبقه‌ها را تشخیص دهد.

### ART۱ -۳

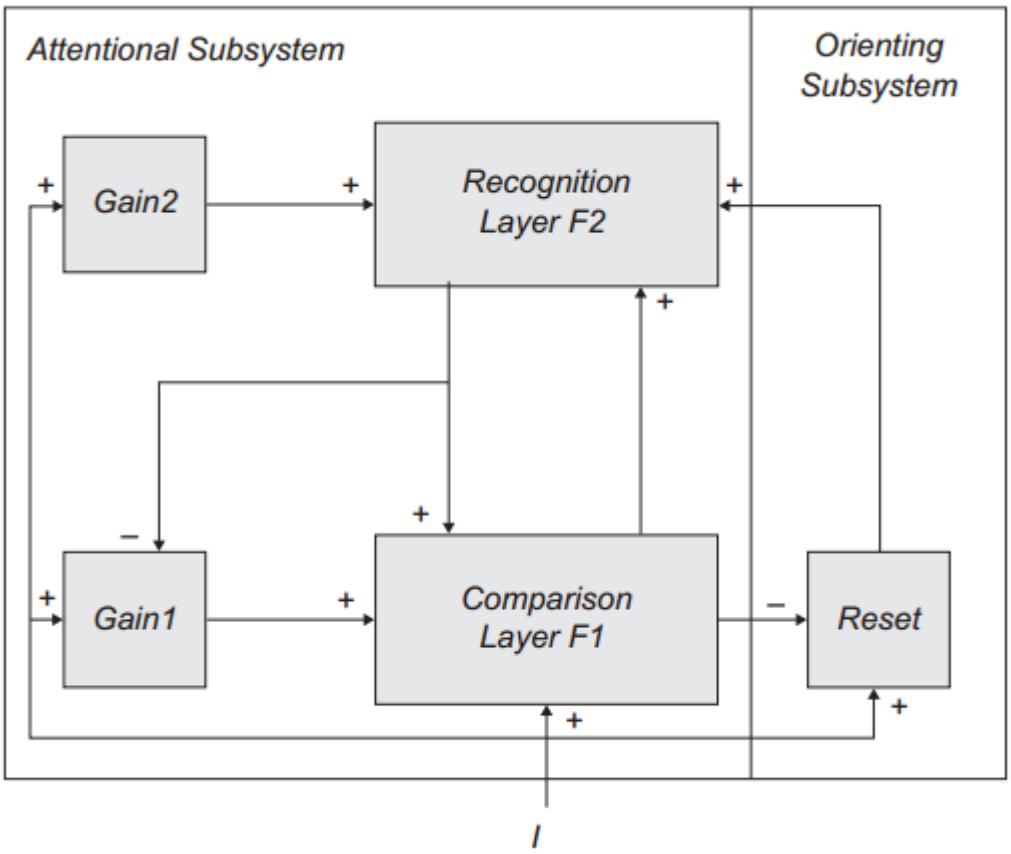
ART۱ ساده ترین مدل یادگیریART می باشد که برای خوشه‌بندی الگوهای دودویی[۱] طراحی شده است. ART۱ شامل دو زیر سیستم توجهی و سوگیری است که در شکل ۱ نشان داده شده‌اند. با تطبیق ورودی پایین به بالا و انتظارات بالا به پایین، یادگیری در زیر سیستم توجهی رخ می‌دهد. زیر سیستم سوگیری هنگامیکه عدم انطباقی در زیرسیستم توجهی رخ دهد، با سیگنال ریست آن را کنترل می‌کند.

زیرسیستم توجهی دو مرحله حافظه کوتاه مدت[۲] (STM) دارد؛ $F_1$ و $F_2$ . $F_1$ حاوی شبکه ای از گره‌ها یا مجموعه ای از سلولهاست که هر کدام نشان دهنده ترکیبی خاص از ویژگی‌های حسی است. یادگیری پایین به بالا، گره‌های $F_2$ را قادر می‌سازد که بصورت انتخابی با ترکیباتی خاص از الگوهای فعالیت در $F_1$ خود را تنظیم کنند و این کار را با تغییر مسیرهای حافظه طولانی مدت[۳] (LTM) بین $F_1$ و $F_2$ انجام می‌دهند. این مسیرها در خروجی سیگنالهای $F_1$ و $F_2$ ضرب می‌شوند. سیگنالهای کنترل بهره[۴] $F_1$ و $F_2$ را قادر می‌سازند که بین مراحل جاری سیکل اجرایی تمایز قائل شوند.

همانطور که در شکل ۱ نشان داده شده است، در مدل ART هر الگوی ورودی خودش را بعنوان یک الگوی فعالیت یا حافظه کوتاه مدت (STM)، در میان ردیاب‌های ویژگی در سطح $F_1$ ثبت می‌کند. هر سیگنال خروجی $F_1$ در وزن‌های انطباقی یا مسیرهای LTM در مسیر مربوطه‌اش ضرب شده و بعد تمام ورودی‌هایی که از گیت LTM گذشته‌اند در گره‌های $F_2$ هدفشان با هم جمع می‌شوند. مهارجانبی یا تعاملات رقابتی در میان گره‌های $F_2$ کنتراس (تضاد) این الگوی ورودی را افزایش می‌دهد. گرچه بسیاری از گره‌های $F_2$ ممکن است ورودی را از $F_1$ دریافت کنند، مهار جانبی موجب می‌شود که مجموعه بسیار کوچکتری از گره‌های $F_2$ فعالیتشان را در STM ذخیره کنند.

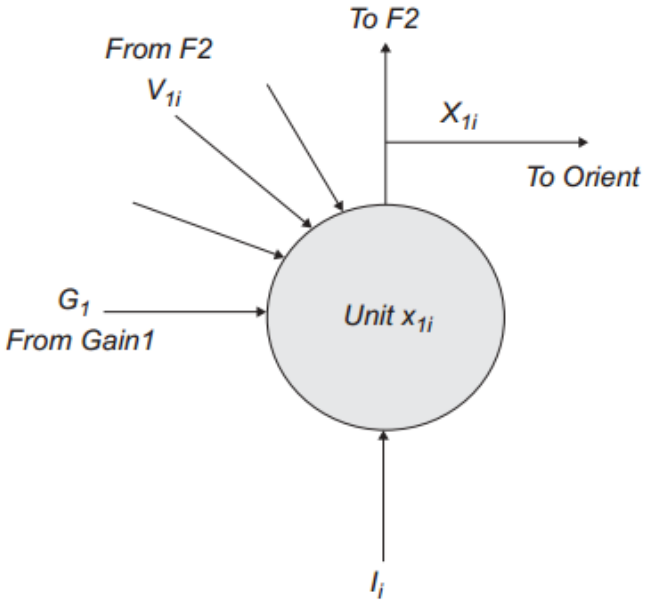
---

[۱] Binary pattern Clustering
[۲] Short-Term Memory
[۳] Long Term Memory
[۴] gain

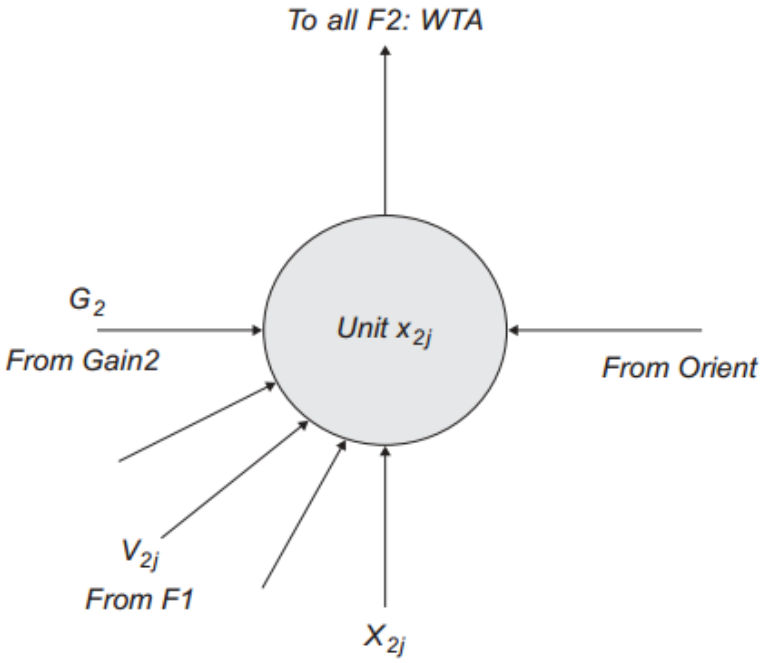



**شکل۱: ART۱ از دوزیر سیستم توجهی وسوگیری تشکیل شده است.**

هنگامیکه عدم تطابقی بین سیگنالهای پائین به بالا و بالا به پائین رخ دهد، سیگنال ریست سلولهای فعال F۲ را مهار می‌کند. ورودی تحریکی به $X_{1i}$ در لایه F۱ از سه منبع وارد می‌شود؛ $V_{1i}$ که خروجی F۲ است ضربدر وزن‌های مربوطه، ورودی خارجی $I_i$ و کنترل بهره $G_1$ و هیچ ورودی مهاری به این نرون وارد نمی‌شود. خروجی نرون به لایه F۲ وارد می‌شود. یک واحد پردازشی $X_{2j}$ در F۲ هم از سه جا ورودی تحریکی می‌گیرد؛ زیر سیستم سوگیری، کنترل بهره $G_2$ ، $V_{2j}$ که خروجی F۱ ضربدر وزن‌های مربوطه است. هیچ ورودی مهاری به آن وارد نمی‌شود. خروجی این نرون به لایه F۱ وارد می‌شود. پردازش در ART۱ به چهار فاز ۱) تشخیص، ۲) مقایسه، ۳) جستجو و ۴) یادگیری تقسیم می‌شود.




**شکل۲: یک واحد پردازنده $X_{1i}$ در لایه $F_1$**




**شکل۳: یک نرون $X_{2j}$ (واحد پردازنده) در $F_2$**

### ۳-۱ فاز تشخیص

در ابتدای فاز تشخیص یا پائین به بالا که هنوز ورودی I اعمال نشده، $G_1$ و $G_2$ صفر هستند و واحدهای $F_2$ غیرفعالند یعنی شانس یکسانی در رقابت تشخیصی بعدی دارند. وقتی ورودی I اعمال می‌شود یک یا چند مولفه‌اش"۱" خواهند بود و این موجب می‌شود $G_1$ و $G_2$ هر دو"۱" شوند. یعنی بهره کنترلی $G_1$ هم به بردار ورودی I و هم به خروجی $X_2$ از $F_2$ بستگی دارد. به بیان دیگر، اگر بردار ورودی I وجود داشته باشد و $F_2$ فعال نباشد (خروجی تولید نکند) $G_1=1$ ،

هرترکیب دیگری از فعالیت I و F۲ موجب مهار $G_1$ می‌شود تا نتواند F۱ را تحریک کند. از طرف دیگر خروجی $G_2$ تنها به بردار ورودی I بستگی دارد:

$$G_1 = \begin{cases} 1 & \text{اگر } X_2 = 0 \text{ و } I \neq 0 \\ 0 & \text{درغیراینصورت} \end{cases} \quad \text{و} \quad G_2 = \begin{cases} 1 & \text{اگر } I \neq 0 \\ 0 & \text{درغیراینصورت} \end{cases} . \qquad (۱)$$

هنگامیکه $G_1 = ۱$، ورودیI وارد F۱ شده $(X = I)$ و سپس واحد تشخیص (F۲) شروع به فعالیت می‌کند. هنگامیکه $G_1 = 0$ ورودی از F۲ وارد F۱ شده و کار مقایسه در F۱ انجام می‌شود. به بیان دیگر اگر بردار ورودی وجود داشته باشد آنگاه $G_2 = 1$ و فرآیند تشخیص در F۲ آغاز می‌شود. هر گرهی در F۱ که یک مقدار ورودی غیرصفر دریافت کند یک الگوی فعالیت STM بزرگتر از صفر تولید و خروجی گره دقیقاً برابر با بردار ورودی می‌شود. از آنجا که $I_i, X_{1i}$ هر دو باینری هستند مقادیرشان"۱" یا صفر است؛ اگر $G_1 = 1$ ، X=I. هر گرهی که فعالیتش بیشتر از حد آستانه باشد یک خروجی تحریکی به گره‌هایF۲ می‌فرستد. الگوی خروجی F۱ یعنی X۱ در وزن‌های $W_{12}$ مسیرهای LTM که F۱ را به F۲ وصل می‌کنند ضرب می‌شود. هر گره F۲ تمام سیگنال‌هایی که از گیت LTM گذشته‌اند را باهم جمع می‌کند؛ $V_{2j} = \sum_i X_{1i} W_{12ji}$. این اتصالات بیانگر دسته‌های طبقه بندی الگوی ورودی هستند، که در آن هر وزن یک طبقه را ذخیره می‌کند. خروجی $X_{2j}$ به این صورت تعریف شده که عنصری که بزرگترین ورودی را دریافت می‌کند باید به روشنی برجسته شود. به این طریق، شبکه رقابتی F۲ به صورت یک شبکه winner-take-all عمل می‌کند؛

$$X_{2j} = \begin{cases} 1 & \text{اگر } G_2 = 1 \text{ و } V_{2j} = \max(V_{2k}) \ \forall_k \\ 0 & \text{در غیر صورت} \end{cases} . \qquad (۲)$$

آن واحد F۲ که بزرگترین خروجی را از F۱ دریافت می‌کند، واحدی است که بهتر از همه با طبقه بردار ورودی منطبق است و در رقابت برنده می‌شود. گره برنده F۲ ، فعال و مقدارش"۱" می‌شود و تمام دیگر گره‌ها را در لایه F۲ مهار کرده و موجب صفر شدن خروجی آن گره‌ها می‌شود.

### ۳-۲ فاز مقایسه

در فاز مقایسه یا انطباق با الگوی بالا به پائین، الگوی فعالیت STM، $X_2$ در F۲ یک الگوی بالا به پائین در F۱ تولید می‌کند. این الگو در $W_{21}$ که F۲ را به F۱ متصل می‌کند ضرب می‌شود. هر گره در F۱ تمام سیگنالهایی که از گیت LTM گذشته‌اند را با هم جمع می‌کند $(V_{1i} = \sum_j X_{2j} W_{21ij})$. فعالترین واحد تشخیصی از F۲ یک"۱" به واحد مقایسه F۱ می‌فرستد و از آنجا که اکنون واحد تشخیص فعال است G۱ مهار و خروجی‌اش صفر می‌شود. طبق قانون "$2/3$" که می‌گوید از سه منبع ورودی مختلف حداقل دوتا باید فعال باشند تا یک خروجی تحریکی تولید شود، تنها واحدهای مقایسه‌ای که فعال می شوند آنهایی هستند که بطور همزمان هم از بردار ورودی و هم لایه تشخیصی مقدار"۱" را دریافت می‌کنند. واحدهایی که یک سیگنال بالا به پائین از F۲ دریافت نمی‌کنند باید غیرفعال باشند حتی اگر ورودی از پائین دریافت کنند؛

$$X_{1i} = \begin{cases} 1 & \text{Ii} \cap \text{V}_{1i=1} \\ 0 & \text{در غیر این صورت} \end{cases} \qquad (۳)$$

اگر انطباق خوبی بین بردار ورودی و قالب[۱] بالا به پائین وجود داشته باشد، سیستم پایدار می‌شود و یادگیری رخ می‌دهد.

[۱] template

اگر عدم انطباقی بین بردار ورودی و خروجی لایه تشخیصی وجود داشته باشد، به این معنی است که الگوی برگردانده شده الگوی مطلوب نیست و لایه تشخیص باید مهار گردد.

### ۳-۳ فاز جستجو

لایه ریست در زیر سیستم سوگیری، میزان شباهت بین بردار ورودی و الگوی خروجی لایه تشخیص را اندازه‌گیری می‌کند. در صورت عدم انطباق بین آنها لایه ریست فعالیت لایه F۲ را مهار می‌کند. زیر سیستم سوگیری بردار ورودی را با خروجی لایه F۲ مقایسه می‌کند و در صورتیکه میزان شباهت کمتر از یک سطح مراقبت باشد، یک سیگنال ریست تولید می‌کند. عدم انطباق با الگوی ورودی در صورتی رخ می‌دهد که $\rho < X_1/I$، ρ در اینجا پارامتر مراقبت ( $0 < \rho \leq 1$ است. اگر تفاوت نسبت دو الگو بیش از مقدار پارامتر مراقبت باشد، یک سیگنال ری ست صادر می‌شود تا واحد فعال کننده در لایه تشخیص F۲ مهار شود. سیگنال ری‌ست موجب می‌شود خروجی لایه تشخیص صفر شود و به مدت زمان طبقه بندی جاری برای جستجو بدنبال یک انطباق بهتر غیرفعال گردد. پارامتر ρ تعیین می‌کند که چقدر عدم تطابق قابل تحمل است. یک پارامتر مراقبت بزرگ موجب می‌شود سیستم در پاسخ به تفاوت‌های کوچک بین I و $X_2$ به دنبال طبقه‌های جدید بگردد و بیاموزد که الگوهای ورودی را به تعدادی زیاد از طبقات ریزتر طبقه بندی نماید. داشتن یک پارامتر مراقبت کوچک اجازه تفاوت‌های زیاد را می‌دهد و الگوهای ورودی بیشتری به یک طبقه یکسان تقسیم می‌شوند.

هنگامیکه یک عدم تطابق رخ می‌دهد، سیگنال مهاری کل از F۱ به زیر سیستم سوگیری افزایش می‌یابد. اگر مهار کافی باشد، زیرسیستم سوگیری فعال می‌شود و یک سیگنال ری‌ست می‌فرستد. سیگنال فعال شده بسته به حالت گره‌های F۲ بر آنها تاثیر می‌گذارد؛ اگر گره F۲ فعال باشد ، سیگنال از طریق مکانیزمی‌که میدان دو قطبی gated خوانده می‌شود موجب یک مهار طولانی مدت می‌گردد. هنگامیکه گره فعال F۲ سرکوب می‌شود، الگوی خروجی بالا به پائین $X_2$ و قالب بالا به پائین $V_1$ حذف می‌شوند و الگوی فعالیت قبلی F۱ مجدداً ایجاد می‌شود. الگوی جدید ایجاد شده $X_1$ موجب می‌شود زیر سیستم سوگیری سیگنال ری‌ست را لغو کند و فعالسازی پائین به بالا دوباره آغاز گردد. از آنجا که گره‌های F۲ که قبلاً فعال شده و دچار مهار طولانی مدت شده‌اند، دراینصورت یک واحد دیگر F۲ در لایه تشخیص برنده می‌شود و یک الگوی ذخیره شده متفاوت به لایه مقایسه بازخورد می‌گردد. اگر الگو دوباره با ورودی منطبق نباشد، کل فرآیند تکرار می‌گردد.

سه مرحله فوق ، یعنی تشخیص، مقایسه و جستجو تا هنگامیکه الگوی ورودی با یک قالب بالا به پایین $X_1$ منطبق گردد، تکرار می‌شوند. در غیر این صورت یک گره F۲ که هنوز هیچ الگویی را نیاموخته است فعال می‌شود. در این حالت گره انتخاب شده F۲ یک طبقه تشخیص الگوی ورودی جدید آموخته شده می‌شود.

### ۳-۴ فاز یادگیری

سه مرحله فوق نسبت به زمان یادگیری مسیرهای LTM بین F۱ و F۲ بسیار سریع رخ می‌دهند در نتیجه می‌توانیم فرض نماییم که یادگیری تنها هنگامی رخ می‌دهد که ری‌ست شدن و فرآیند جستجو خاتمه می‌یابد و تمام الگوها در F۱ و F۲ پایدار می‌شوند. مسیرهای LTM بین F۱ و F۲ از معادله زیر پیروی می‌کند:

$$\tau_1 \frac{dW_{12ij}}{dt} = \begin{cases} (1 - W_{12ij})L - W_{12ij}(|X_1| - 1) & if\ V_{1i} \text{ and } V_{1j} \text{ are active} \\ -|X_1|W_{12ij} & if\ only\ V_{1j} \text{ is active} \\ 0 & if \text{ only } V_{1j} \text{is inactive} \end{cases} \quad (٤)$$

که در آن $\tau_1$ ثابت زمانی و $L$ پارامتری با مقدار بزرگتر از یک است. از آنجا که $\tau$ به اندازه کافی نسبت به زمان فعالسازی $STM$ بزرگتر و از زمان ارائه الگوی ورودی کوچکتر است، معادله فوق یک معادله با سرعت یادگیری کم است که به معادله با سرعت یادگیری بالای زیر همگرا می‌شود:

$$W_{12ij} = \begin{cases} \dfrac{L}{L-1+|X_1|} & if\ \mathrm{V_{1i}}\ \text{and}\ \mathrm{V_{1j}}\ \text{are active} \\ 0 & if\ only\ \mathrm{V_{1j}}\ \text{is active} \\ \text{no change} & if\ \text{only}\ \mathrm{V_{1j}}\ \text{is inactive} \end{cases} \quad (۵)$$

مقدار اولیه $W_{12ij}$ باید بصورت تصادفی در بازه زیر انتخاب شود:

(٦)

$$0 < W12ij < \frac{L}{L-1+|M|}$$

که در آن $M$ بعد الگوی ورودی برابر با تعداد گرهها در $F_1$ است. مسیرهای $LTM$ بین $F_2$ و $F_1$ از معادله زیر پیروی می‌کنند:

$$\tau_2 \frac{dW_{21ji}}{dt} = X_{2j}\left(-W_{21ji} + X_{1i}\right) \quad (۷)$$

که در آن $\tau_2$ ثابت زمانی است و معادله فوق در حین ارائه یک الگوی ورودی به معادله با سرعت یادگیری بالای زیر همگرا می‌شود:

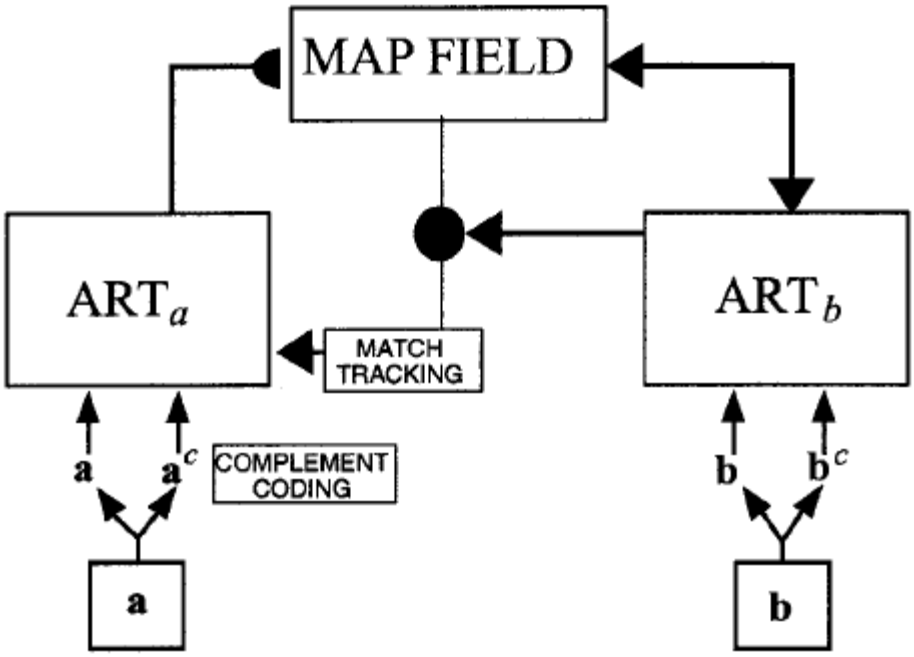
$$W_{21ji} = \begin{cases} 1 & if\ \mathrm{V_{1i}}\ \text{and}\ \mathrm{V_{1j}}\ \text{are active} \\ 0 & if\ \text{only}\ \mathrm{V_{1i}}\ \text{is inactive} \end{cases} \quad (۸)$$

که در آن مقدار اولیه $W_{21ji}$ بصورت تصادفی مقداری در بازه مقابل انتخاب می‌شود: $1 \geq W_{21ji}(0) > C$

که در آن برحسب پارامترهای معادله یادگیری کُند، در مورد C تصمیم گیری می‌شود ، هرچند مقدار تمام $W_{21ji}(0)$ها در حالت یادگیری سریع"۱" انتخاب می‌شود.

## ۴- یادگیری با سرپرست و پیشگویی در ARTMAP

یک سیستم ARTMAP حاوی یک جفت ماژول ART است، $ART_a$ و ، $ART_b$ (شکل۴). در طی یادگیری با سرپرست، $ART_a$ یک دنباله از الگوهای $\{a^{(n)}\}$ را دریافت می‌کند و $ART_b$ یک دنباله از الگوهای $\{b^{(n)}\}$ را که در آن $b^{(n)}$ پیشگویی صحیحی است با فرض $a^{(n)}$ (شکل۴).




**شکل۴: شبکه کلی ARTMAP برای یادگیری با سرپرست شامل دو ماژول ART . برای تکالیف طبقه بندی $ART_b$ می‌تواند ساده شود.**

یک شبکه یادگیری تداعیگر و یک کنترلر پارامتر مراقبت این ماژولها را لینک می‌کند تا موجب شود سیستم ARTMAP بصورت بلادرنگ کار کند، و حداقل تعداد طبقه‌های تشخیص $ART_a$ ، یا واحدهای پنهان لازم برای برآوردن معیار دقت را ایجاد کند. یک قانون minimiax، ARTMAP را قادر می‌سازد تا به سرعت، به دقت و بصورت کارآمد بیاموزد، زیرا آن بصورت برخط و بطور تلفیقی خطای پیشگویی را حداقل می‌کند و فشردگی کد را حداکثر می‌کند. یک پارامتر $\overline{\rho_a}$ مراقبت خط پایه حداقل معیار انطباق را تنظیم می‌نماید. $\overline{\rho_a}$ کوچکتر اجازه تشکیل طبقه‌های بزرگتر را می‌دهد. در ابتدای یک آزمایه تعلیم $\rho_a = \overline{\rho_a}$ یک خطای پیشگویی در $ART_b$، $\rho_a$ را درست به اندازه کافی افزایش می‌دهد تا یک جستجو را آغاز کند، از طریق یک مکانیزم کنترل فیدبک که ردیابی انطباق خوانده می‌شود. یک کد فعال جدید توجه را بر یک خوشه متفاوت از ویژگی‌های ورودی متمرکز می‌کند و چک می‌کند که آیا این ویژگیها بهتر قادر به پیشگویی خروجی صحیح هستند. ردیابی انطباق به ARTMAP اجازه می‌دهد که یک پیشگویی را برای یک رخداد نادر بیاموزد، که در انبوهی از رخدادهای مکرر مشابه که پیشگویی متفاوتی می‌کنند نهفته شده است.

ARTMAP یک مرحله پیش پردازش را بکار می‌برد که کدگذاری مکمل خوانده می‌شود. این مرحله با نرمال کردن الگوهای ورودی، مساله بالقوه ازدیاد طبقه‌ها را حل می‌کند (۶). کدگذاری مکمل تعداد مولفه‌های ورودی را دو برابر می‌کند؛ به شبکه هم بردار، ویژگی اولیه و هم مکمل آن را ارائه می‌نماید. به بیان نوروبیولوژیکی، کدگذاری مکمل برای بازنمایی یک الگوی ورودی هم از مرکز-روشن[۱] و هم از اطراف-خاموش[۲] استفاده می‌کند. بخش مرکز-روشن مربوطه از یک بردار وزن ویژگی‌هایی را رمز می‌کند که بطور مداوم غالب هستند. وزن‌های کوچک در بخش‌های مکمل یک بازنمایی طبقه بعنوان انطباق بی ارزش کدگذاری می‌شوند؛ آن ویژگی‌هایی که گاهی حاضر و گاهی غایب هستند.

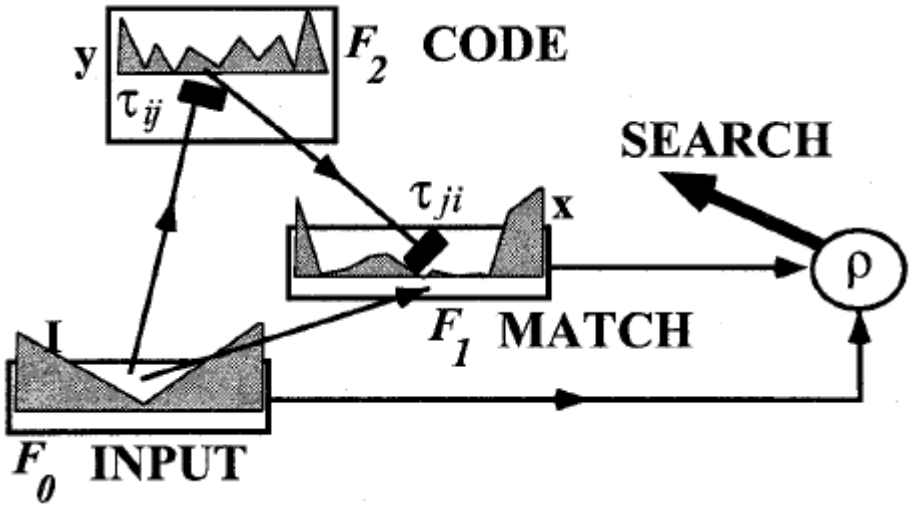
---

[۱] on-cell
[۲] off-cell

### ۵- کدگذاری توزیع شده در dART

فعالیت winner-take-all در شبکه‌های ART از کدگذاری پایدار پشتیبانی می‌کند اما هنگامیکه ورودی‌های نویزدار با یادگیری سریع تعلیم داده می‌شوند موجب ازدیاد طبقه‌ها می‌شود. در مقابل، فعالیت مک کولوخ-پیتز[۱] توزیع شده در MLPها تحمل نویز را بهبود می‌بخشد. اما یادگیری سریع موجب فراموشی مصیبت بار می‌شود. مدل‌های ART توزیع شده[۲] (dART) هدفشان پل زدن بین این دو دنیا است: فعالیت توزیع شده تحمل نویز را افزایش می‌دهد، در حالیکه دینامیک‌های جدید سیستم قابلیت‌های یادگیری پایدار سیستم‌های ART winner-take-all را حفظ می‌کند(۱۸). این شبکه‌ها بصورت اتوماتیک تغییرات آموخته شده را بر طبق درجه فعالیت هر گره کدگذار تقسیم می‌کنند .این امر موجب یادگیری توزیع شده سریع و همچنین کُند می‌شود بدون اینکه فراموشی فاجعه باری رخ دهد.

قوانین جدید یادگیری انتقال سیناپسی در شبکه dART دوباره پیکره بندی شده، از مشکلات محاسباتی که هنگام اعمال کدگذاری توزیع شده روی معماری یک شبکه ART سنتی رخ می‌دهد گریز می‌کند (شکل۵). عنصر طراحی اصلی که اجازه می‌دهد dART مشکل فراموشی فاجعه بار یادگیری توزیع شده سریع را حل کند وزن پویا است. این کمیت معادل با تفاوت جبران شده بین فعالیت گره کدگذار و یک مقدار آستانه انطباقی است که به این وسیله حافظه کوتاه مدت و بلند مدت را در واحد اصلی محاسباتی شبکه ترکیب می‌نماید.




**شکل۵: معماری توزیع شده ART (dART) که پایداری شبکه‌های ART را حفظ می‌کند ولی اجازه می‌دهد کد $F_2$ درمیان نودهای دلخواه توزیع شود.**

مقدار آستانه $Z_{ij}$ در مسیرهایی که مستقیماً یک میدان ورودی $F_0$ را به میدان کدگذاری $F_2$ تصویر می‌کنند از قانون یادگیری سن[۳] توزیع شده پیروی می‌کند (dInstar)که هنگامیکه کد گذاری winner-take-all است به قانون سن فروکاهش می‌یابد.

یادگیری در مسیرهای $F_0 \rightarrow F_2$ بجای بهره انطباقی شبیه میزان تأثیر سیناپسی دوباره توزیع شده[۴] (RSE) است که توسط مارکرام و تسودیکس[۱] (۱۹) در سیناپس‌های قشری مشاهده شد. در این آزمایش‌ها جفت شدن، قوت یا تاثیر انتقال

---

[۱] Mcculloch-Pitts
[۲] Distributed ART
[۳] Instar
[۴] Redistribution of Synaptic Efficacy

سیناپسی را برای ورودی‌های تست با فرکانس پائین افزایش می‌دهد اما برای ورودی‌های تست با فرکانس بالا نمی‌تواند قوت سیناپسی را افزایش دهد و حتی ممکن است قوت سیناپسی را کم کند. در سیستم یادگیری dART ، RSE دقیقاً دینامیک محاسباتی لازم برای پشتیبانی از کدگذاری توزیع شده پایدار بلادرنگ است.

مقدارهای آستانه $Z_{ji}$ در مسیرهایی که میدان کدگذار F۲ را به یک میزان انطباق دهنده F۱ تصویر می‌کنند از قانون outstar توزیع شده (dOutstar) پیروی می‌کنند که اصل لاغری ناشی از ترک (عدم استعمال) یادگیری انتظارات شبکه با توجه به الگوی فعالسازی میدان کدگذاری توزیع شده را تحقق می‌بخشد. مثل سیستم‌های ART winner-take-all ، dART انتظارات بالا به پایین را با ورودی پائین به بالا در میدان انطباق دهنده مقایسه می‌کند و اگر انطباق نتواند معیار پارامتر مراقبت را برآورده کند، به سرعت به دنبال یک کد جدید می‌گردد.

## ۶- نتیجه گیری

در این مقاله، نظریه تشدید انطباقی (ART)، برخی کاربردهای آن به اجمال، اصول کارکرد آن و همچنین اجزای تشکیل دهنده سه مدل ART یعنی ART[۱]، ARTMAP و dART به تفصیل، ارائه شد.ART مساله پایداری-انعطاف پذیری را حل می کند، زیرا می‌تواند بسرعت و بصورت پایدار بیاموزد بدون اینکه آنچه که قبلا آموخته است را بطور فاجعه باری فراموش کند. با تغییر اندازه پارامتر مراقبت ART، یک سیستم می‌تواند هم طبقه‌های انتزاعی مثل صورت‌ها و سگ‌ها و هم نمونه‌های منفرد از این طبقه‌ها را تشخیص دهد. دو زیر سیستم توجهی و سوگیری در ART[۱] ، بعلاوه مراحل پردازشی در آن شامل ۴ مرحله تشخیص، مقایسه، جستجو و یادگیری با جزئیات تشریح شد.

## ۷- مراجع

۱. [illegible], [illegible]. (۱۹۷۶[illegible]) [illegible] [illegible] [illegible] [illegible] [illegible]: [illegible]. [illegible] development and coding of neural feature detectors, Biological cybernetics, ۲۳(٤),۱۲۱–۱۳٤.

۲. [illegible], [illegible]. (۱۹۷۶[illegible]). [illegible] [illegible] [illegible] [illegible] [illegible]: [illegible]. [illegible], expectation, olfaction, illusions. Biological cybernetics, ۲۳(٤), ۱۸۷-۲۰۲.

۳. [illegible], [illegible]. [illegible]. (۲۰۱۸). [illegible] [illegible] [illegible] [illegible] [illegible]. [illegible].

۴. [illegible], [illegible]. [illegible]., [illegible], [illegible]. [illegible]. (۲۰۰۶). [illegible] [illegible]. [illegible] [illegible] [illegible] [illegible]. [illegible] ۱۰,۱۱,۲.

۵. [illegible], [illegible]. [illegible]., [illegible], [illegible]. (۱۹۸۷). [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]-[illegible] neural pattern recognition machine. Computer vision, graphics, and image processing, ۳۷(۱), ٥٤-۱۱٥.

۶. [illegible], [illegible]. [illegible]., [illegible], [illegible]., [illegible], [illegible]. [illegible]. (۱۹۹۱). [illegible] [illegible]: [illegible] [illegible] [illegible] [illegible] categorization of analog patterns by an adaptive resonance system. Neural networks, ٤(٦), ۷٥۹-۷۷۱.

۷. [illegible], [illegible]. [illegible]., [illegible], [illegible]., [illegible], [illegible]., [illegible], [illegible]. [illegible]., [illegible], [illegible]. [illegible]. (۱۹۹۲). [illegible] [illegible]: [illegible] network architecture for incremental supervised learning of analog multidimensional maps. IEEE Transactions on neural networks, ۳(٥), ٦۹۸-۷۱۳.

[۱] Markram & Tsodyks

۸. □□□□□□□□□□, □. (۱۹۷۸). □□□□□□□□□□□□□□□, □□□□□□□, □□□□ □□□□□□□□□□□□□. □□□□□□□□□ □□ □□□□□□□□□□□□□ □□□□□□□□, ۵, ۱۸۳-۲۳۲.

۹. Von der Malsburg, C. (۱۹۷۳). Self-organization of orientation sensitive cells in the striate cortex. Kybernetik, ۱٤(۲), ۸۵-۱۰۰.

۱۰. □□□□□□□□□, □. □., □ □□□□ □□□□ □□□□□□□□□□, □. (۱۹۷۶). □□□□ □□□□□□□□□□ □□□□□□□ □□□□□□□□□□□□□ □□□□ □□ □□□□ □□ □□ self-organization. Proceedings of the Royal Society of London. Series B. Biological Sciences, ۱۹٤(۱۱۱۷), ۴۳۱-۴۴۵.

۱۱. □□□□□□□□, □. □., □□□□□□, □. □., □□□□□□□□□□, □., □ □□□□□□□□□□, □. (۱۹۹۴). □□□□□: □□□□□□ □□□□□□ □□□□-۱ □□□□□□□ architectures for engineering design retrieval. Neural Networks, ۷(۹), ۱۳۳۹-۱۳۵۰.

۱۲. □□□□, □. □., □□□□□□, □. □., □□□□□□, □. □., □□□□□□, □. □., □□□□□□□□□□, □. □., □ □□□□□□□, □. □. (۲۰۰۱, □□□□□□□). □□□□□□□ □□ multisensor passive and active ۳D imagery. In Enhanced and Synthetic Vision ۲۰۰۱ (Vol. ٤۳٦۳, pp. ۲۱۹-۲۳۰). □□□□□.

۱۳. *□□□□□□□□□, □., □□□□□, □. □., □ □□□□□, □. □. (□□□□.). (۱۹۹۹). □□□□□□□□ □□□□□□□□□: □□□□□□□□□ □□□□□□□□□□□□ perception and performance. MIT Press.*

۱۴. □□□□□□□□-□□□□□□□□□□□□□, □., □□□□□□□□-□□□□□□□□□, □., □ □□□□□□□□, □. □. (۲۰۱۲). □□□□□□□□□ □□□□□□□□□□ □□□□□□□ microchips: circuit design techniques (Vol. ٤٥٦). Springer Science & Business Media.

۱۵. □□□□□□, □., □ □□□□□□□, □. (۱۹۸۹). □□□□□□□ □□□□□□□□□ □□□□ □□□□□□□□□□ □□□□□□□□□□ □□□□□□□□□: □□□□□□□□□ □□□□□ examples without local minima. Neural networks, ۲(۱), ۵۳-۵۸.

۱۶. □□□□□, □., □ □□□□□, □. (۱۹۹۲). □□ □□□□ □□□□□□□□ □□ □□□□□□ □□□□□□□ □□ □□□□□□□□□□□□□□□□□. □□□□□ □□□□□□□□□□□□□□ on Pattern Analysis and Machine Intelligence, ۱٤(۱), ۷٦-۸٦.

۱۷. □□□□□□□□□□, □. (۱۹۸۲). □□□□ □□□□□ □ □□□□□□ □□□□□□ □ □□□□□□□□□□ □□□□□□. □□ □□□□□□□□ □□ □□□□□ □□□□ □□□□□□ (□□. ۱-۵۲). □□□□□□□□□, □□□□□□□□□□.

۱۸. □□□□□□□□□□, □. □. (۱۹۹۷). □□□□□□□□□□□□ □□□□□□□□□, □□□□□□□□□□□□, □□□□ □□□□□□□□□□□ □□ □□□□ □□□□ □□□□□□□ neural networks. Neural networks, ۱۰(۸), ۱٤۷۳-۱٤۹٤.

۱۹. □□□□□□□□, □., □ □□□□□□□□, □. (۱۹۹۶). □□□□□□□□□□□□□□□□ □□ □□□□□□□□□ □□□□□□□□□ □□□□□□□□ □□□□□□□□□□□□ pyramidal neurons. Nature, ۳۸۲(٦۵۹٤), ۸۰۷-۸۱۰.