فصلنامه علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

سال هفتم، شماره‌های ۲۵ و ۲۶، پاییز و زمستان ۱۳۹۴
صص: ۴۳- ۶۶

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

# تولید نام‌های زیبا و معنادار ایرانی به‌کمک الگوریتم ژنتیک با تابع برازندگی مبتنی بر شبکهٔ عصبی مصنوعی

[*]امیرشهاب شاهمیری [**]بهاره زمانی نظامی [***]سعید شیری
*دانشجو دکتری، مدیریت آموزش عالی، دانشگاه آزاد اسلامی، واحد علوم تحقیقات، تهران
** دانشجوی کارشناسی ارشد، دانشکدهٔ فنی و مهندسی، دانشگاه آزاد اسلامی واحد پرند، تهران
*** استادیار، دانشکدهٔ مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران
تاریخ دریافت: ۹۳/۰۱/۲۵ تاریخ پذیرش: ۹۴/۱۱/۲۰

**چکیده**

زبان‌های گفتاری جهان در طول هزاران سال تاریخ خود، روندی تکاملی را پیموده و به امروز رسیده‌اند. بر همین راه واژگان زبان‌های گوناگون و البته نام‌های متداول در فرهنگ‌ها و زبان‌های سراسر جهان نیز، به تکامل رسیده‌اند و این روند تکاملی را ادامه خواهند داد. برخی از این نام‌ها، با توجه به حس زیبایی‌شناسی افراد یک جامعه و ویژگی‌های آوایی زبانی، که بسته به زمان و مکان متغیر است، رواج بیشتری می‌یابند و تطور و گسترش بیشتری پیدا می‌کنند.

در این پژوهش با شناخت واژگان و نام‌های فارسی، مدلی پیشنهاد شده که بر پایهٔ آن نام‌هایی نو، که همراه با داشتن معنایی درخور، از نظر زیبایی‌شناسی نیز نظر افراد گوناگون را برآورده سازد، تولید شود. ساخت واژگان تازه به‌کمک پردازش تکاملی (الگوریتم ژنتیک) انجام می‌شود که در آن، تابع برازندگی را (به‌جز در نسل نخست) یک شبکه عصبی مصنوعی از نوع پیشخور پس‌انتشار با دو لایه پنهان که تا ۸۹٪ مشابه انسان تخمین می‌زند، شبیه‌سازی می‌کند. بهترین نام تولید شده با این روش، رتبه ۵ را در میان ۲۵۲ نام جامعه اولیه به‌دست آورده است. بهترین نمونه‌های جامعهٔ اولیه و داده‌های آموزشی و آزمایشی و نیز برآورد تابع برازندگی نسل نخست و سنجش زیبایی بهترین نام‌های تولید شده بدین روش، با تحلیل آماری نمونه‌های اخذ شده به‌کمک پرسش‌نامه انجام شده است.

**واژه‌های کلیدی:** فرهنگ نام‌های ایرانی، هوش مصنوعی، واژه‌شناسی، نام‌شناسی، پرسپترون چندلایه.

## ۱.مقدمه

نام‌ها مهم‌ترین عامل برای بازشناسی اشخاص، اشیا و مفاهیم در فکر ما هستند. ما محیط پیرامون خود را به‌یاری نام‌ها می‌شناسیم و درک می‌کنیم. نام‌هایی عام که مفاهیم کلی چون درخت، فلز، جانور و غیره را در ذهن ما تداعی می‌کنند و نام‌های خاص که موجب دسته‌بندی و شناسایی موجودات و رویدادها در یاد ما می‌شوند.

روشن است که لفظی آهنگین و خوش‌نوا با مفهومی زیبا و پسندیده که بیش از هر واژه دیگری طی زندگانی شخص تکرار و در گوشش طنین‌افکن می‌شود، القای مثبتی را در پی دارد و بر منش و رفتار فرد تاثیر خواهد گذاشت. پژوهش در فرهنگ نام‌گزینی مردم ایران نشان می‌دهد که آنان در گزینش نام نیک برای فرزندان خود نیز از ذوق هنرمندانه و حس زیباشناسی بی‌همتایی برخوردار هستند و ایرانیان بر

نویسنده عهده‌دار مکاتبات:امیرشهاب شاهمیری amir@shahmiri.ir

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

خوش‌آوایی و گوش‌نوازی نام‌ها نیز تکیه می‌کنند[۱].
طبق نظرسنجی که طی فرایند انجام همین پژوهش صورت گرفته، جامعهٔ آماری هدف، هنگام گزینش نام برای فرزندشان، نزدیک به ۳۰٪ به خوش‌آوایی و ۷۰٪ معنای نام اهمیت می‌دهند.
اما نام‌های ایرانی نیز همچون خود فرهنگ ایرانی، در طول تاریخ دستخوش دگرگونی، پیشرفت و تکامل شده‌اند تا به شیوهٔ امروزی درآمده‌اند و بر شناسنامه ایرانیان می‌نشینند. برای نمونه، نام/واژهٔ وهومننگهه (vohumanangha) به‌معنای (دارندهٔ) اندیشهٔ نیک در زبان اوستایی به وهومن (vohuman) در پهلوی تبدیل شده و امروزه به‌شکل هومن و بهمن به‌کار می‌رود. یا ارتخشثره (artaxshathra) اوستایی، به‌معنای پادشاهی/ شهر مقدس، به ارتخشر (artaxsher) پهلوی تبدیل شده و امروزه به‌شکل اردشیر خوانده می‌شود. با همین روند ان‌اهیته (an-ahita) اوستایی به‌معنای پاک (ضد ناپاک) به آناهیتا، آناهید و ناهید امروزی تبدیل شده است [۲و۳]. همچنین طی هزاران سال، با بهره‌گیری از گنجینهٔ واژگان ایرانی و وندهای آن، انبوهی از نام‌های ترکیبی زیبا و بامعنا ساخته و در اختیار مردمان ایران و جهان قرار گرفته است، نام‌هایی مانند بهنام، مهسا، شیرین، مهرنوش و فرزاد.
این روند دگرگونی و تولید نام‌های تازه در زبان‌های ایرانی هرگز از حرکت باز نایستاده و در آینده نیز متوقف نخواهد شد و از آن‌جاکه روند تکاملی آن از قواعد خاص مرتبط با زبان‌شناسی و دستور، و نیز آواشناسی واژگان ایرانی، پیروی می‌کند و همچنین از آن‌جا که معیار و مقدار زیبایی نام‌های کهن و نوپدید توسط ایرانیان را به‌سادگی می‌توان از تعداد گزینش این نام‌ها توسط ایشان بازشناخت، به‌نظر می‌رسد که با بهره‌گیری از برخی روش‌های هوش مصنوعی، مانند پردازش تکاملی[۱] و الگوریتم ژنتیک[۲] و نیز آگاهی کافی از دانش زبان‌شناسی و دگرگونی تاریخی واژگان، می‌توان این روند تکاملی را شبیه‌سازی کرده، به واژگان جدیدی دست یافت.
در ادامهٔ این بخش، پس از بیان اهداف پروژه، پیشینهٔ پژوهش‌ها و کارهای انجام شدهٔ مرتبط با این پروژه بررسی می‌شود. سپس در بخش دوم، ویژگی‌های واژگان و نام‌های فارسی و شیوهٔ واژه‌سازی و روش آمیزش و ترکیب واژگان با یکدیگر برای ساخت نام‌ها در این زبان به‌کوتاهی معرفی خواهد شد. بخش سوم مدل طراحی‌شده برای انجام این پژوهش را نمایش خواهد داد و در بخش چهارم روش کار و اجرای پروژه، مشتمل بر تشکیل پایگاه دادهٔ نام‌ها و چگونگی استخراج ویژگی از آن و کارکرد الگوریتم ژنتیک و شبکهٔ عصبی مصنوعی[۳] به‌کار رفته در این پروژه و نیز تحلیل و ارزیابی نتایج کار شرح داده خواهد شد. در فصل پنجم جمع‌بندی و پیشنهاد کارهای آینده ارایه می‌شود. دو پرسش‌نامه از نام‌های اولیه و تولیدی نیز پیوست شده است.

### ۱-۱- هدف پژوهش

نخستین هدف این پژوهش، آزمون و سنجش کارآمدی روش‌ها و ابزارهای هوش مصنوعی برای انجام پروژه‌ها و حل مسایل علوم انسانی و به‌ویژه زبان‌شناسی و ادبیات بوده است. به‌انجام رسیدن این پروژه ما را بیشتر متقاعد می‌کند که تحلیل و یافتن پاسخ مسایل پیچیدهٔ علوم گوناگون، به‌کمک روش‌های نوین هوشمند امکان‌پذیر است. دیگر هدف این پروژه نیز تولید نام‌های زیبا و معنادار در زبان فارسی و ارایه آن به جامعه برای به‌کارگیری در نام‌گذاری افراد، شرکت‌ها و مواردی از این دست بوده است.

### ۱-۲- پیشینهٔ پژوهش

تاکنون پژوهش‌های عملی و نظری بسیاری برای پردازش زبان فارسی به‌کمک روش‌های هوش مصنوعی انجام پذیرفته است. برای نمونه، شاهمیری و همکاران، در سال‌های ۱۳۸۴ تا ۱۳۸۷ سرایندگان اشعار فارسی را به‌کمک روش‌های یادگیری ماشین با دقت درستی ۹۴ تا ۱۰۰٪ شناسایی و تعیین کردند[۴و۵]، با شبکه‌های عصبی مصنوعی توانسته‌اند ۸۰ تا ۱۰۰٪ از غلط‌های تایپی در متن فارسی را تصحیح کنند[۶] و نیز ریشهٔ زبانی واژگان فارسی را از عربی با دقت ۹۲٪ تشخیص دهند[۷]. در سال ۱۳۸۸، اصفهانی و همکاران سیستمی را برای شناسایی و طبقه‌بندی نام افراد، مکان‌ها، سازمان‌ها و ... ارایه و برای آموزش آن یک بردار ویژگی استخراج کردند و به میانگین دقت ۹۶٪ دست یافتند[۸]. آرمین و شمس‌فرد در سال ۱۳۸۹ الگوریتمی را برای ریشهٔ واژگان و تبدیل واژگان محاوره‌ای به رسمی با

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

به‌کارگیری روش ان‌گرام‌ها پیشنهاد کردند[۹]. در سال ۱۳۹۱، ریاحی و همکاران از الگوریتم ژنتیک و شبکهٔ عصبی RBF برای تهیهٔ یک سیستم خلاصه‌سازی متن فارسی استفاده کردند[۱۰]. بابادی و همکاران برای یافتن ریشهٔ واژگان فارسی، به‌همراه پیشوندها و پسوندهای آن، قواعد استنتاج و شبکهٔ عصبی RBF را به‌کار بردند[۱۱]. زاهدی و همکاران برخی از روش‌های ریشه‌یابی واژگان فارسی را (مانند ریشه‌یاب‌های آماری، مبتنی بر ساختار و چندفازه)، برای شناخت تاثیر آنها بر سیستم‌های بازیابی اطلاعات متنی بررسی کردند[۱۲].

از سوی دیگر، تاکنون پژوهش‌ها وکتاب‌های بسیاری در زمینهٔ نام‌های ایرانی به‌انجام رسیده است. برای نمونه، پورپارسی، آوا و معنای واژگان نام‌های ایرانی از پارسی باستان، اوستایی و پهلوی دوره‌های مادی، هخامنشی، اشکانی و ساسانی را بررسی کرد[۱۳]. حمیدا نیز آوانگاری، ریشه زبانی، معنا و جنسیت نزدیک به ۱۰ هزار نام ایرانی را ارایه کرد[۱]. زاهدی‌نیا و کاظمی، آوانگاری، ریشه زبانی، معنا و جنسیت نام‌های ایرانی را در دو لیست جداگانه، با نمایش ریشه‌های فارسی، عربی، ترکی، کردی، لاتین، پهلوی، اوستایی، مازندرانی، یونانی، عبری، سانسکریت، سغدی، فارسی، ترکمنی، ارمنی و لری، گرد آوردند[۱۴]. همچنین مبشر آوانگاری، ریشهٔ زبانی، معنا و جنسیت نام‌ها و نیز اسامی اسطوره‌ای و اسامی ایرانی، عربی و قومیت‌های گوناگون را برشمرد[۱۵]. زنگنه نیز آوانگاری، ریشه زبانی، معنا و جنسیت نام‌های ایرانی را بررسی کرد[۱۶]. انصاری و افتخار، آوانگاری، ریشهٔ زبانی، معنا و جنسیت نام‌های ایرانی و همچنین نام‌هایی که می‌توانند به‌عنوان نام خانوادگی یا نام شرکت‌های تجاری به‌کار روند را تعیین کردند[۱۷]. بهنیا نیز ریشهٔ زبانی، معنا و جنسیت هر نام را به‌همراه فراوانی هر یک در میان تعداد کل مردان یا زنان و همچنین مجموع فراوانی نام‌هایی که با هر یک از حروف الفبا آغاز می‌شوند، ارایه کرد که البته، تعداد نام‌ها کم، محدود و بسیار قدیمی است و نام‌های امروزی و جدید را در بر ندارد[۱۸]. اکبری بیرقی هم دائره‌المعارف نام‌های ایرانی را به‌همراه آوانگاری، ریشهٔ زبانی و معنای واژگان را ارایه کرده است[۱۹]. به‌جز منابعی که ذکر شد، فرهنگ معین، در بخش اعلام (جلد ۵ و ۶) [۲۰] و لغت‌نامهٔ دهخدا [۲۱] نیز به نام‌های ایرانی پرداخته‌اند، که نام‌های به‌کار گرفته شده در این پژوهش، با آنها نیز مقرانه شده است.

## ۲- ویژگی نام‌ها و واژگان ایرانی

هر چیز مادی یا مفهومی با نام آن شناخته می‌شود. نام‌ها نشانگر همهٔ مفاهیم شناخته شده برای ما، فارغ از زمان و مکان یا وابسته به آن هستند. اما نام‌ها و واژگان هر زبان، به آن زبان وابسته‌اند. پس برای ایجاد نام در یک زبان باید چگونگی و ساختار نام و واژه را در آن زبان شناخت. فرهنگ بریتانیکا نام‌ها را در ۹ طبقه جای می‌دهد[۳۴]:

۱) نام‌های شخصی
۲) نام‌های شبه‌شخصی (مانند نام حیوانات)
۳) نام جای‌ها یا مکان‌ها
۴) نام‌هایی که به اشیای بدون تشخص انسانی داده شده، اما هم‌چنان برجسته و چشم‌گیر است (مانند سرزمین مقدس)
۵) نام‌های گروه‌های اجتماعی و قومیت‌ها و سازمان‌ها
۶) نام آثار هنری
۷) نام برندها یا کالاهای تجاری
۸) نام حوادث یا دوره‌های تاریخی
۹) نام‌های ذهنی و تجریدی که تعیین عینی یا تشخص مشخصی ندارند، مانند مشاییان، بی‌دینان.

یک نام ممکن است هم‌زمان در دو یا چند طبقه از طبقات یاد شده قرار گیرد.

نام‌های شخصی همان شیوهٔ تکاملی و انتقالی و تطوری را می‌پیمایند که دیگر اجزای زبانی و واژه‌های روزمره طی می‌کنند. آنها ممکن است به‌تدریج از یک فرهنگ، جذب فرهنگی دیگر شوند؛ چنان‌که انبوه نام‌های ایرانی به زبان‌های ترکی، عربی و اروپایی راه یافته و از این زبان‌ها نیز نام‌های بسیار پذیرفته است.

پیدایش نام‌های ترکیب‌شده با پسوند در فرهنگ ایرانی نیز پدیده‌ای به‌نسبت نوین است و از آن‌جاکه سطح دانش و فهم زبانی مردم افزایش یافته، پیشوندها یا پسوندهای خوش‌معنا و زیبا همچون شکیب، مهر، آریا و جم رایج شده است. گفتنی است که وقتی سخن از نام‌های ایرانی می‌شود، منظور

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

نام‌هایی است که در گسترهٔ فرهنگی ایران شکل گرفته و نهاده می‌شود که این حوزه فرهنگی به‌مراتب و گسترده‌تر از حوزهٔ جغرافیایی ایران است[۱۵].

### ۲-۱- نام و معنای آن

دانشی که به بررسی نام‌ها و وجوه آن می‌پردازد، نام‌شناسی[۴] یا علم‌الاعلام خوانده می‌شود[۳۴]. موضوع این دانش بسیار گسترده است زیرا حوزهٔ مطالعاتی آن به همهٔ عرصه‌هایی که به نام و نیز به قلمرویی که به‌طور نظری در برگیرندهٔ همهٔ زبان‌ها، عرصه‌های جغرافیایی و فرهنگی و عصرهای تاریخی است، مربوط می‌شود. برای مقاصد اجرایی، برخی تقسیم‌بندی‌های موضوعی (از نظر زبانی، جغرافیایی، تاریخی و ...) ضرورت دارد.

امروزه یکی از مهم‌ترین عناصر در گزینش نام، توجه به‌معنای نام و مفاهیم مرتبط با آن است و این امر ناشی از افزایش آگاهی‌های زبانی و کسب دانش زبان ادبی است. ورود نام‌هایی با معانی زیبا به قلمرو نام‌های ایرانی، مانند فرزانه به‌معنای صاحب خرد و اندیشه، ترانه به‌معنای کلام آهنگین، تیراژه به‌معنای رنگین‌کمان و ده‌ها نام خوش‌معنای دیگر پی‌آمد همین آگاهی است.

گزینش نام فرزندان توسط ایرانیان به چند عامل بستگی داشته است، از جمله: پایگاه فرهنگی، اجتماعی و قومی والدین و نیز جنسیت فرزند[۱۵].

لازم به ذکر است، فرهنگ و زبان‌های ایرانی، بر خلاف بیشتر زبان‌ها و فرهنگ‌های دیگر، از دورهٔ باستان، از دیدگاه فرهنگی و اجتماعی تفاوتی میان مرد و زن قایل نمی‌شده و با وجود آن‌که در زبان‌های بسیار کهن ایرانی نیز مانند بسیاری از زبان‌های امروز جهان (مانند فرانسه، عربی، اسپانیایی و...) جنسیت در واژگان نقش داشته و نام‌ها و برخی کلمات مذکر یا مونث بوده‌اند[۲۲]، این تمایز کم‌کم در زبان‌های ایرانی از میان رفته است و امروزه هم بسیاری نام‌های ایرانی را هم برای دختران و هم برای پسران برمی‌گزینند و تنها نرمی یا سختی کلامی یا معنایی و یا گاه تاریخچهٔ آن نام (مانند مه‌رو، نرگس و آزاده برای دختران و گرشاسپ، شیراوژن و رستم برای پسران) تا حدودی جنسیت آن را تعیین می‌کند. از این‌رو در این پروژه، جنسیت نام‌ها در نظر گرفته نشده است.

در این‌جا باید خاطرنشان کرد که در زبان‌ها و گویش‌های ایرانی (کردی، گیلکی، خراسانی و ...) کمتر نامی بدون مفهوم و معنای روشن وجود دارد و اگر در بررسی‌های ادبی به نام‌هایی برمی‌خوریم که به‌ظاهر دارای معنا نیستند، ممکن است مفهوم آنها در گذر تاریخ گم‌شده و از یاد رفته باشد[۱۵]. در این پژوهش نیز نام‌هایی برای درج در پایگاه دادهٔ نام‌ها برگزیده شده‌اند که از معنای آنها اطمینان کافی وجود داشت.

### ۲-۲- نام از نظر ساختاری

تحول و تکامل زبانی نام‌ها امری پیچیده است. در نظام زبانی هندواروپایی باستان هر فرد یک نام داشت که ممکن بود به دو صورت باشد: مرکب یا بسیط. نام‌های بسیط اصولا به کسانی داده می‌شد که اعضای فرودست قبیله بودند. بهره‌گیری از نام‌های مرکب در ایران امروز امری نامتداول نیست، اما ایرانیان امروزی نام‌های ساده خوش‌صوت با موسیقی کلامی و نیز خوش‌معنا را ترجیح می‌دهند. با این‌حال تعداد نام‌های مرکب اندک نیست و ساختار دستوری نام‌های مرکب بسیار متنوع است، از جمله[۱۵]:

- ترکیب دو اسم: روان‌مهر، آریامهر، آذرسام، آذرکیش، آذرنوش.
- ترکیب اسم و صفت: مهرافزا، گلشن‌آرا، پاریاب، جهان‌بخش.
- ترکیب اسم و حرف: افسانه (ه: ابرازساز)، چوبک (ک تصغیر)، فریبا (آ: صفت مشبه (اسم‌ساز)).
- ترکیب اسم و صفت مفعولی: شهداد، کوه‌زاد، فرادید.
- ترکیب اسم و یای نسبت: فرهی، رهی.
- ترکیب اسم و پسوند: الوند، فرازمند.

### ۲-۳- ریشه زبانی نام‌های ایرانی

یک زبان از دو بخش اصلی واژگان و دستورزبان تشکیل می‌شود. زبان فارسی یکی از زبان‌های آریایی[۵] یا هندواروپایی[۶]، است آریایی واژه‌ای است که آن را پژوهشگران آلمانی و ایرانی به‌کار می‌برند؛ زیرا منشا آن از اقوام آریایی ساکن در ایران بوده است و هندواروپایی را

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

زبان‌شناسان انگلیسی (به پیشنهاد فرانتس بوپ[۷] در سال ۱۸۱۶) ترجیح می‌دهند، زیرا گستره این زبان‌ها از هندوستان تا اروپا را در بر گرفته است. البته برخی زبان‌شناسان آلمانی ادعا دارند که این زبان‌ها از هند تا ایسلند، که زبانی ژرمنی است، کاربرد دارند و از این‌رو باید آنها را هندوژرمنی[۸] خواند و برخی نیز آریایی را برابر با شاخهٔ هندوایرانی می‌دانند[۲۳].

### ۲-۴- زبان‌های آریانی

آریانی یا هندواروپایی به خانواده بزرگی از زبان‌ها گفته می‌شود. این خانواده شامل تقریبا همهٔ زبان‌های فلات ایران، شبه‌قارهٔ هند و اروپا می‌شود. در پانصد سال گذشته کوچ اروپاییان به قاره‌های آمریکا، استرالیا و آفریقا، دامنه این زبان‌ها را به آن‌جا نیز گسترش داده است. البته در گذشته زبان مردم آسیای صغیر (آذربایجان، ترکیه و گرجستان)، آسیای مرکزی (ترکمنستان، ازبکستان، و ...) و چین غربی (ایالت سین‌کیانگ) نیز آریانی بوده اما ورود قبایل ترک‌زبان از حدود ۱۵۰۰ تا ۷۰۰ سال پیش، موجب از میان رفتن زبان‌های محلی آن‌جا (مانند آذری و تاتی) شده است[۲۴]. شکل ۱، طبقه‌بندی زبان‌های آریانی امروز و زبان‌های ایرانی را نشان می‌دهد.

### ۲-۴- زبان‌های ایرانی

زبان‌های شناخته‌شدهٔ ایرانی را با توجه به معیارهای اولیهٔ برون‌زبانی ، به‌طور قراردادی به سه دوره تقسیم کرده‌اند که عبارتند از: ایرانی باستان (تا سده‌های چهارم و سوم پ.م)، ایرانی میانه (تا سده‌های هشتم و نهم م.) و ایرانی نو (تا امروز) [۲۵].

البته در زنجیرهٔ دگرگونی و پیشرفت زبان‌ها از آریایی تا فارسی امروز، زبان فارسی روندی همچو زنجیره شکل ۲ را پیموده است.

هرچند در این زنجیره، طبقه‌بندی زبان‌های آریایی تا ایرانی آغازین مورد توافق همه زبان‌شناسان نیست و برخی هر سه را در حقیقت یکی می‌دانند که دیگر زبان‌های آریایی نیز از آن ریشه می‌گیرند. اما با توجه به اسناد موجود و کتیبه‌ها و نبشته‌های باستانی، در مورد زبان‌های ایرانی باستان، میانه و نو اتفاق نظر وجود دارد.

زبان اوستایی به‌همراه فارسی باستان تنها زبان‌های ایرانی باستان هستند که از خود نوشته به‌جای گذاشته‌اند. فارسی باستان در اسناد سیاسی زمان حکومت هخامنشیان به‌کار رفته و در حقیقت نشان‌دهنده دست‌کم یک گویش از گویش‌های ناحیه پارس و یکی از گویش‌های جنوب‌غربی ایران است. اوستایی نیز در کتاب مقدس ایرانیان باستان، یعنی اوستا، و برخی متون دیگر، موجود و با زبان سنسکریت بسیار همانند است.

زبان‌های ایرانی نو هم با فروپاشی شاهنشاهی ساسانی شکل گرفتند. فارسی دری (مادر فارسی امروزی)، کردی، بلوچی، پشتو، آسی، تاتی، تاجیکی، از مهم‌ترین زبان‌های ایرانی نو هستند[۲۵].

در طول زمان، واژگان فارسی، همچون واژگان دیگر زبان‌های جهان، دگرگون شده‌اند و به تکامل رسیده‌اند و البته به بسیاری از دیگر زبان‌های جهان راه یافته‌اند.

جدول ۱ روند دگرگونی برخی از واژگان ایرانی و رسوخ آن در برخی از زبان‌ها را نشان می‌دهد.

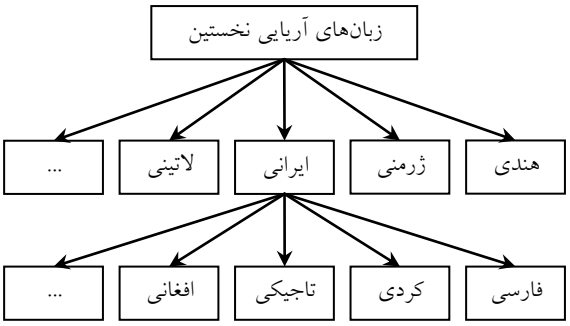
### ۲-۵- واژه‌سازی در زبان فارسی

بزرگ‌ترین واحد هر زبان، جملهٔ مستقل و کوچک‌ترین آن واج است. واحدهای زبان از کوچک‌ترین تا بزرگ‌ترین دارای سلسله مراتب زیر است: واج (واژ یا حرف)، تکواژ (واژک)، واژه، گروه، جمله و جملهٔ مستقل [۲۶] که البته در این پژوهش تنها به سه عنصر نخست توجه می‌شود.

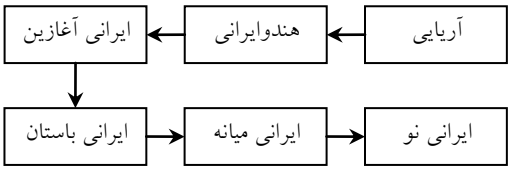
صرف یا ساخت‌واژه[۹] بخشی از دستور است که چگونگی ساخت واژه در زبان مورد نظر را تحلیل می‌کند[۲۷].

**واج:** کوچک‌ترین واحد سخن است که تنها صوت دارد، اما معنا ندارد و در عین حال سبب تمایز معنایی می‌شود. تعداد واج‌ها در فارسی ۲۳ همخوان (صامت) و ۶ واکه (مصوت) است [۲۸]. در این پژوهش (برای هماهنگی با نمادهای به‌کار گرفته شده در برنامه‌نویسی الگوریتم ژنتیک و شبکه عصبی مصنوعی) در آوانگاری از حروف ساده لاتین استفاده شده و نه نمادهای استاندارد جهانی (جدول ۴ را ببینید).

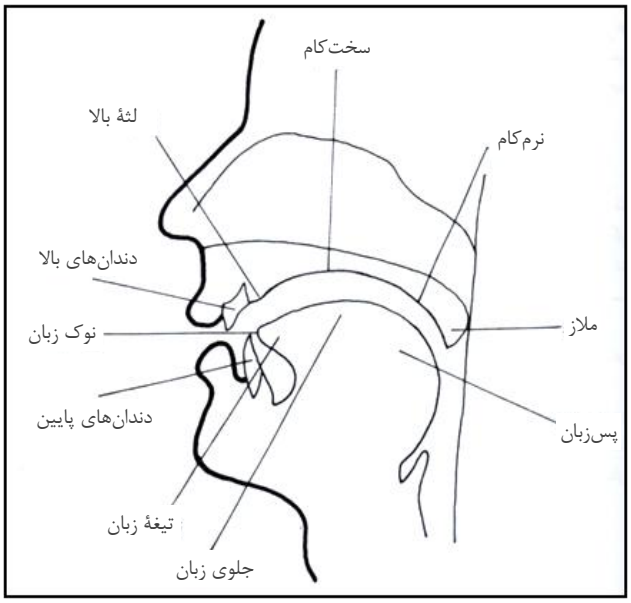
Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026




**شکل ۱) طبقه‌بندی زبان‌های آریایی [۲۴]**




**شکل ۲) روند تحول زبان‌های ایران تا به امروز [۲۵]**




**شکل ۳) بخش‌های گوناگون دهان و زبان در رابطه با ساخت آوا [۲۷].**

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

**جدول ۱) برخی نام‌های ایرانی و مشتقات آنها در زبان‌های دیگر[۳و۲۴]. (آوانگاری بر پایهٔ جدول ۵)**

| زبان | واژه | | | | |
|---|---|---|---|---|---|
| **فارسی‌امروزی** | خسرو | مهر / میترا | ستاره / اختر | پری | پردیس |
| **فارسی باستان** | hao-srawngha | miθra | stArem | pairika | pairi-daeza |
| **اوستایی** | hu-sravah | miθra | stArem/staro | pairikA | pairi-daeza |
| **سانسکریت** | suSravas | mitrah/mitrA | naksatra/tAras | - | - |
| **پهلوی** | husruv/xusraw | mitr/mihr | stArak/star | parik / parig | pairi-daeza |
| **لاتین** | caesar | mitre | astrum/stella | pretium | paradisus |
| **یونانی** | kaisar | mitrhe | aster | pallakis | paradeisos |
| **آلمانی** | kaiser | mithra | stern | fee | papadies |
| **انگلیسی** | caesar | mithras | star | fairy / fay | paradise |
| **معرب** | کسری/ قیصر | مهر | اصطر(لاب) | - | فردوس |

از دیدگاه محل تولید آواهای زبان و برخورد آن با دیگر عضوهای دهان، زبان را می‌توان به شش بخش تقسیم کرد[۲۹]:

۱) نوک زبان که پشت دندان‌های پایین جای می‌گیرد.
۲) تیغه زبان که زیر ناحیه لثه بالا جای می‌گیرد.
۳) جلوی زبان که زیر ناحیه سخت‌کام جای می‌گیرد.
۴) مرکز زبان که زیر ناحیه مرکزی سخت‌کام جای می‌گیرد.
۵) پشت زبان که زیر نرم‌کام جای می‌گیرد.
۶) ریشه زبان که زیر ملاز جای دارد و تا دیوارهٔ جلویی حلق ادامه می‌یابد.

شکل ۳ ساختمان زبان در دهان را نمایش می‌دهد. شناخت ساختار سیستم تولید آوا در دهان کمک می‌کند تا چگونگی تغییر و تبدیل واج‌های تشکیل‌دهندهٔ واژگان زبان‌ها را بهتر بشناسیم..

**تکواژ:** از یک یا چند واج ساخته می‌شود و کوچک‌ترین واحد معنادار در زبان است؛ مانند: من مداد خر ید م.

**واژه:** از یک یا چند تکواژ ساخته می‌شود و استقلال نحوی، جدایی‌پذیری و یکپارچگی دارد. گل، گلاب، گلاب‌گیر و گلاب‌پاش همگی واژه‌ای مستقل هستند[۲۶].

واژه را می‌توان از چهار دیدگاه تعریف کرد:

۱) **آوایی:** واژه یک ساخت آوایی است که از چند هجا تشکیل شده، دارای یک تکیه است و در آغاز و پایان آن یک درنگ (سکوت) وجود دارد.

۲) **ساخت صرفی:** واژه از یک یا چند واژک تشکیل شده و در سلسله مراتب واحدهای دستوری زبان در ساختمان گروه به‌کار می‌رود.

۳) **معنایی:** واژه عبارتست از یک واحد معنایی که بر یک یا چند مفهوم منفرد دلالت دارد.

۴) **نگارشی:** دارای وحدت املایی است، یعنی پیش و پس از آن فاصله گذاشته می‌شود [۲۸].

در این پژوهش به‌ترتیب، دیدگاه‌های معنایی و ساخت صرفی و آوایی بیشتر مدنظر قرار گرفته‌اند.

همچنین می‌توان واژه را به دو گونهٔ ساده و غیرساده تقسیم کرد. واژهٔ ساده تنها یک تکواژ و واژهٔ غیرساده بیش از یک تکواژ دارد[۲۸]. در این پروژه تنها نام‌های ۲ و ۳ تکواژی برای درج در پایگاه دادهٔ نام‌ها برگزیده شده‌اند.

گفتنی است که واژه در زبان فارسی از ترکیب تکواژها و وندها و دیگر اجزا ساخته می‌شود و بدین ترتیب با وجود محدودیت تکواژها، تعداد واژگان زبان فارسی بی‌نهایت است. اما برای مثال، زبان عربی به‌هیچ وجه چنین توانی را ندارد و ریشهٔ واژگان ثلاثی و رباعی آن در قالب وزن‌های محدودی صرف و واژگان محدودی ساخته می‌شود[۳۰].

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

**جدول ۲) گردش‌های مجاز برخی واکه‌های فارسی به‌کار گرفته شده در این پژوهش به‌همراه نمونه[۲،۳و۲۰].**

| واکه‌ها (مصوت‌ها) | اَ (a) | آ (A) | اِ (e) | ایـ (i) | اُ (o) | او (u) | حذف واکهٔ آغاز یا آخر |
|---|---|---|---|---|---|---|---|
| اَ (a) | × | پَد: پاد | وَهیشت: بِهشت | ارتَوهیشت: اردی‌بهشت | اهورامَزدا: هرمُز درَگ: دروغ | | انوشیروان: نوشیروان |
| آ (A) | ماه: مَه | × | کارمانیا: کِرمان | – | – | – | آناهیتا: ناهید |
| اِ (e) | اِراک: اَراک | – | × | انوشه‌روان: انوشیروان | وِمان: گُمان | – | اسپید: سپید |
| ایـ (i) | اژی‌دهاک: ضَحاک | – | وهیشت: بهِشت | × | بیرون: بُرون | – | سامی: سام |
| اُ (o) | وُهومن: بَهمن | – | – | – | × | رُچ: روز | – |
| او (u) | بورزاَخَو: بَرزخ | – | – | – | اهورامزدا: هُرمز | × | مغو: مغ |
| ـه (a/e) پایانی | فرزانَه: فرزانِه | انوشه: انوشا | – | – | – | – | فرزانه: فرزان |

**جدول ۳) دسته‌بندی همخوان‌های فارسی بر پایهٔ شیوه و محل تولید آوا در دهان[۲۷].**

| محل تولید / نحوه تولید | دولبی | لب و دندانی | دندانی | لثوی | لثوی و کامی | کامی | ملازی | چاکنایی |
|---|---|---|---|---|---|---|---|---|
| انفجاری | ب/پ | | ت/د | | | ک/گ | غ | آ/ئ/ع* |
| سایشی | | و/ف | | س/ز | ش/ژ | | خ | ه |
| انفجاری/سایشی | | | | | ج/چ | | | |
| لرزشی | | | ر | | | | | |
| روان | | | ل | | | ی | | |
| غنه‌ای | م | | ن | | | | | |

### ۲-۶- جابه‌جایی واکه‌ها در زبان فارسی

گردش واکه‌ها (مصوت‌ها)[۱۰] از ویژگی‌های زبان‌های آریایی است که توسط یاکوب گریم[۱۱] در سال ۱۸۲۲ معرفی شد. گردش واکه‌ها عبارت از جابه‌جا شدن مصوت‌ها در ریشه‌ها، پسوندها، پیشوندها، میانوندها و اجزای صرفی واژگان است. اگر مصوت‌های کوتاه و بلند مانند a و A جابه‌جا شوند (مانند ماه و مه)، گردش را گردش کمّی و اگر مصوت‌های گوناگون مانند a، e و o جابه‌جا شوند (مانند مادر و mother)، این گردش را گردش کیفی می‌گویند[۲۹].

در زبان‌های هندوایرانی در زمینهٔ آوایی سه گونه دگرگونی ویژه رخ می‌دهد که آنها را از دیگر زبان‌های هندواروپایی متمایز می‌سازد [۲۳]:

۱) تغییرات و تفاوت‌های کیفی (باز یا بسته و پیشین یا پسین بودن واکه)

۲) جایگزینی واکه میانی خنثی در هجای فاقد تکیه (ə)

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

با واکه نامعین یا کسره کوتاه (e)

۳) دگرگونی گروه واج‌های س (s) هندواروپایی به گروه واج‌های ش (š) هندوایرانی، در صورتی‌که پس از e و o و همخوان‌های کناری و پسکامی درآیند.

جدول ۲ برخی از گردش‌های مجاز در واکه‌های زبان‌های ایرانی را (که در این پژوهش نیز به‌عنوان دگرگونی مجاز برای جهش ژنتیکی در الگوریتم ژنتیک به‌کار رفته است) نمایش می‌دهد. یادآور می‌شود که شمار واکه‌ها و نیز دایرهٔ جابه‌جایی‌های آنها، در طول تاریخ زبان‌های ایرانی، بسیار فراتر از این موارد بوده است. در زبان فارسی همهٔ همخوان‌ها، به‌جز چند آوا (مانند: نچ)، برون‌سو هستند و بسته به شیوه و محل تولید آوا در دهان دسته‌بندی می‌شوند. جدول ۳ این تقسیم‌بندی را نشان می‌دهد.

همخوان‌های موجود در زبان فارسی نیز مانند واکه‌های آن، در طول زمان بسیار به یکدیگر تبدیل شده‌اند[۱۵و۲۵]. جدول ۳ در بخش جهش ژنتیکی این پژوهش به‌کار گرفته شده؛ به‌این ترتیب که درصورت وقوع جهش که بر پایه یک تابع تصادفی تعیین می‌شود، حروف داخل هر خانه با احتمال ۲ برابر نسبت به حروف سطر یا ستون همسایه بالا، پایین، چپ یا راست و ۴ برابر همسایه‌های مورب خود به یکدیگر تبدیل می‌شوند. برای نمونه:

$$P(\text{گ}\rightarrow\text{ک}) = 2P(\text{غ}\rightarrow\text{ک}) = 4P(\text{ش}\rightarrow\text{ک})$$

گفتنی است که در زبان ایرانیان گاهی واکه‌ها و همخوان‌ها به یکدیگر تبدیل شده‌اند (مانند مَیَزد به میز و خسرَوْ به خسرو) اما در این پروژه از این‌گونه تبدیلات صرف‌نظر شده است.

## ۳- مدل‌سازی

برای پیاده‌سازی و به‌انجام رسانیدن این پروژه، مدلی طراحی شد که سه فاز اصلی را در بر دارد. در فاز نخست و سوم، کار به‌صورت پژوهش میدانی و تحلیل اطلاعات، و در فاز دوم، به‌صورت شبیه‌سازی کامپیوتری انجام شده است. (ر.ک. بخش ۴)

همان‌گونه که در شکل ۴ دیده می‌شود، پس از به‌دست آوردن دانش مورد نیاز برای شناخت واژگان و نام‌های ایرانی، در فاز نخست پایگاه داده‌ای نام تشکیل و با توجه به نظر افراد جامعهٔ آماری و برخی ملاحظات زبان‌شناسانه هرس می‌گردد. ویژگی‌هایی که بتوان به آنها مقادیر کمّی اختصاص داد نیز در این فاز از نام‌ها برکشیده می‌شود.

سپس در فاز دوم همزمان با تنظیم و آموزش شبکهٔ عصبی، نخستین نسل از نام‌ها تولید خواهد شد که برازندگی این نام‌ها و نام‌های نسل‌های بعدی را همان شبکهٔ عصبی تعیین می‌کند.

پس از تولید نام‌ها به‌تعداد کافی دلخواه، در فاز سوم بهترین نام‌های تولیدشده به همان افراد جامعهٔ آماری اولیه برای رتبه‌بندی واگذار شد تا درستی کارکرد روش سنجیده، و تحلیل و ارزیابی مناسب انجام پذیرد.

## ۴- روش کار

در این بخش به شرح عملکرد بخش‌های تعریف‌شده در مدل شکل ۴ می‌پردازیم.

### ۴-۱- ساخت پایگاه نام‌های آغازین

برای ساخت و تکمیل پایگاه داده‌ای از نام‌های ایرانی، با بررسی کتاب‌های بسیار (که در بخش پیشینه پژوهش بدان‌ها اشاره شده است) مرجع [۱] به‌عنوان منبع اصلی و دیگر منابع برای مقارنه و تقابل برگزیده شدند.

از این مجموعه، ۵۰۰ نام به‌شیوه‌ای نیمه‌تصادفی از کتاب انتخاب شد. نیمه‌تصادفی از این جهت که از هر برگ کتاب یک یا دو نام برگزیده شد؛ مشروط بر این‌که برخی شرایط ابتدایی مورد نیاز پژوهش به‌شرح زیر را دارا باشد:

۱) نام ریشهٔ ایرانی داشته باشد.

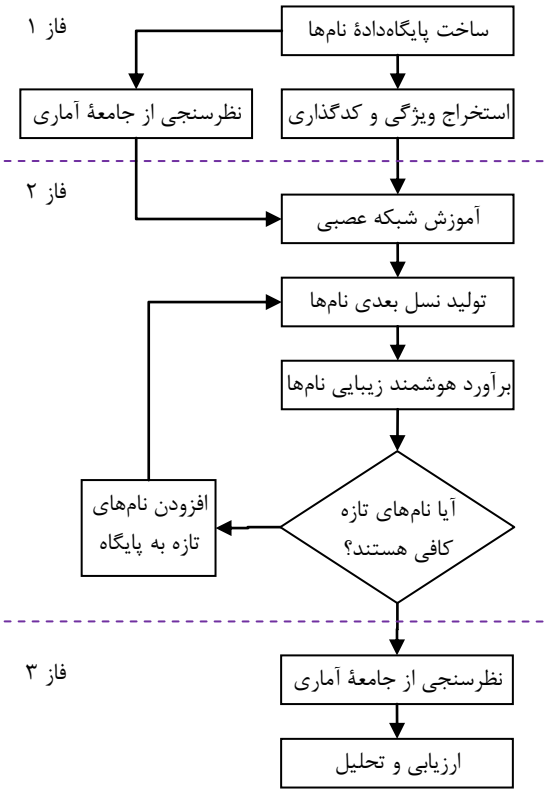
۲) نام فقط از ۲ تکواژ فارسی ترکیب شده باشد، مانند مهتاب: مه + تاب.

۳) نام دارای ۲ یا ۳ هجا (سیلاب) باشد، مانند فرزانه: فر + زا + نه.

۴) نام دارای معنای روشن و مشخص باشد.

پس از گردآوری نام‌های انتخاب شده با شرایط بالا، با مراجعه به سایت سازمان ثبت احوال کشور (شهریور ۱۳۹۰)، تعداد ثبت هر یک از نام‌ها در این سازمان، که به‌نوعی نشان‌دهندهٔ میزان تمایل یا استقبال مردم از این نام‌هاست، در پایگاه نام‌ها درج شد.

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026




**شکل ۴) مدل ارایه شده برای تولید نام**

سپس نام‌ها بر اساس فیلد فراوانی نام به‌صورت صعودی مرتب شدند و رتبهٔ آنها به‌عنوان میزان برازندگی آنان در نظر گرفته شد و بر این اساس، همان‌گونه که در جدول ۴ نمایش داده شده است، با یک تابع ساده، ۲۵۲ نام به‌صورت تصادفی انتخاب گردید. بدین ترتیب نام‌های با رتبهٔ پایین‌تر احتمال انتخاب بالاتر داشته‌اند (مینا بالاترین و راوک پایین‌ترین احتمال انتخاب).

نام‌های برگزیده، هم به‌عنوان اعضای جمعیت اولیه در الگوریتم ژنتیک و هم به‌عنوان ورودی آموزشی شبکهٔ عصبی برای به‌دست آوردن تابع برازندگی الگوریتم به‌کار خواهند رفت.

### ۴-۲- تهیهٔ پرسش‌نامه و نظرسنجی از جامعهٔ آماری

از آن‌جاکه سیستم ارایه‌شده در این پروژه باید بتواند نظر زیبایی‌شناسانهٔ مردم را نسبت به نام‌ها شبیه‌سازی کند، لازم است که یا پارامترهای زیبایی‌شناسی ایرانیان نسبت به این نام‌ها استخراج، تدوین و در پروژه به‌صورت محاسبه‌پذیر اعمال شود؛ که البته این کار، کاری بسیار دشوار است و پژوهشی چندجانبه را می‌طلبد، و یا این‌که چکیدهٔ نظر آنان به‌روشی انتزاعی در پروژه منظور شود.بدین منظور، ۲۵۲ نام برگزیده، در قالب ۶ فرم، هر یک با ۴۲ نام تقسیم شد و فرم‌ها در طول سال ۹۰ در جامعه آماری ویژه‌ای (حدود ۳۰۰ تن، اغلب از میان دانشجویان، استادان و شاغلان دانشگاه آزاد اسلامی واحد پرند) توزیع و پس از تکمیل، گردآوری شد (نمونهٔ پرسش‌نامه در پیوست۱). ). از این‌رو روشن است که تابع برازندگی بیشتر نظرات این جامعهٔ آماری را تخمین خواهد زد. از آن‌جاکه زیبایی از دو دیدگاه معنایی و آوایی بر انتخاب یک نام در اندیشهٔ ایرانیان تاثیر می‌گذارد، از جامعهٔ آماری درخواست شد تا در دو ستون جداگانه به زیبایی آوایی و معنایی هر یک از نام‌ها نمره‌ای بین صفر تا ۱۰ را اختصاص دهند.

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

**جدول ۴) نام‌های استخراج شده از پایگاه نام‌های اولیه، به همراه فراوانی آنها در پایگاه وب سازمان ثبت احوال کشور و رتبهٔ آن**

| ردیف | نام | فراوانی | ردیف | نام | فراوانی |
|---|---|---|---|---|---|
| ۱ | راوک | ۲۷ | ⋮ | ⋮ | ⋮ |
| ۲ | ایرمان | ۲۹ | ۲۴۴ | بهرام | ۱۱۷۴۱۲ |
| ۳ | تیما | ۳۵ | ۲۴۵ | مژگان | ۱۲۲۶۲۶ |
| ۴ | کامکار | ۳۸ | ۲۴۶ | فرهاد | ۱۶۵۷۸۵ |
| ۵ | تیراژه | ۴۱ | ۲۴۷ | پریسا | ۱۶۶۰۰۴ |
| ۶ | رایمند | ۴۵ | ۲۴۸ | افسانه | ۱۷۰۴۵۲ |
| ۷ | بهناد | ۵۲ | ۲۴۹ | امید | ۲۰۵۶۲۴ |
| ۸ | آپامه | ۵۷ | ۲۵۰ | فرشته | ۲۰۶۳۵۳ |
| ۹ | راتا | ۵۹ | ۲۵۱ | فرزانه | ۲۱۶۳۸۲ |
| ⋮ | ⋮ | ⋮ | ۲۵۲ | مینا | ۲۱۸۶۵۲ |

همچنین از ایشان درخواست شد تا درجهٔ اهمیت زیبایی آوایی و معنایی را در هنگام گزینش نام به‌درصد مشخص کنند.

میانگین نمرات تخصیص‌یافته، به‌عنوان بردار هدف برای آموزش شبکهٔ عصبی و همچنین به‌عنوان میزان شایستگی در تابع برازندگی الگوریتم ژنتیک (نسل یکم) به‌کار گرفته می‌شود.

### ۴-۳- استخراج ویژگی

جدا از طراحی و تنظیم تابع ارزیابی، مهم‌ترین مساله در الگوریتم ژنتیک، تدوین نقشهٔ ژن‌ها و چگونگی تولید نسل تازه است.

در الگوریتم ژنتیک ژن‌ها باید در ساختاری استاندارد قرار گیرند و محل گسست دو ژن و تولید نمونهٔ جدید[۱۲] به‌درستی تعیین شود[۳۲].

از این‌رو لازم آمد تا هر واژه آوانگاری، و نیز محل‌های مجاز برای انجام عمل گسست و ترکیب ژن‌ها تعیین شود. تعیین درست مکان گسست و ترکیب ژن‌ها، معنادار بودن واژهٔ تولیدشده را تضمین می‌کند.

جدول ۵ به‌ترتیب آوانگاری (فونتیک) معادل هر یک از واکه‌ها (مصوت‌ها) حروف همخوان (صامت) زبان فارسی به‌کار گرفته شده در این پروژه را نمایش می‌دهد.

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

**جدول ۵) آوانگاری واکه‌ها و همخوان‌های زبان فارسی**

| ردیف | آوا | نماد | نمونه | نگارش |
|---|---|---|---|---|
| ۱ | اَ | a | بد | bad |
| ۲ | اِ | e | دل | del |
| ۳ | اُ | o | بت | bot |
| ۴ | ـا | A | پا | pA |
| ۵ | ای | i | سی | Si |
| ۶ | او | u | بو | bu |
| ۷ | اَو / اُو | w | روشن | rwSan |
| ۸ | آ/ او/ ایـ/ ئـ/ ع | & | آرشیدا | &ArSidA |
| ۹ | ب | b | بانو | bAnu |
| ۱۰ | پ | p | پوری | puri |
| ۱۱ | ت (ط) | t | تارا | tArA |
| ۱۲ | ج | j | جاوید | jAvid |
| ۱۳ | چ | c | چاوش | cAvoS |
| ۱۴ | خ | x | خسرو | xosrw |
| ۱۵ | د | d | دانا | dAnA |
| ۱۶ | ر | r | رامبد | rAmbod |
| ۱۷ | ز (ذ) | z | زیبا | zibA |
| ۱۸ | ژ | Z | ژاله | Zale |
| ۱۹ | س (ث) | s (θ) | سمن | saman |
| ۲۰ | ش | S | شیوا | SivA |
| ۲۱ | غ (ق) | q | قباد | qobAd |
| ۲۲ | ف | f | فرزین | farzin |
| ۲۳ | ک | k | کارن | kAren |
| ۲۴ | گ | g | گلناز | golnAz |
| ۲۵ | ل | l | لاله | lAle |
| ۲۶ | م | m | مینا | mina |
| ۲۷ | ن | n | نرگس | narges |
| ۲۸ | و | v | وارَش | vAraS |
| ۲۹ | ه | h | هامون | hAmun |
| ۳۰ | ی | y | یکتا | yektA |

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

البته از آن‌جاکه آوای حروف ح، ص، ض، ط، ظ، ع و ق در زبان فارسی وجود ندارند، در این جدول نادیده گرفته شده‌اند و حروف ث و ذ نیز امروزه به‌صورت س و ز خوانده می‌شوند.

البته در آوانگاری این پروژه، مانند روش معمول زبان‌شناسی فارسی، از واکهٔ w به‌عنوان یک واج واحد چشم‌پوشی شده است.

بدین ترتیب پایگاه دادهٔ نام‌ها به‌همراه آوای آنها به‌تفکیک هجای نام مربوطه و همچنین به‌تفکیک واژگان مجرد و وندهای تشکیل‌دهندهٔ واژه از نظر معنایی، ساخته شد. جدول ۶ برخی از واژگان را در این پایگاه داده نمایش می‌دهد. در این جدول، فیلد بار که بار منفی معنای بخش پیشوندی نام را نشان می‌دهد، برای این درج شده که هنگام آمیزش دو نام، معنای منفی یا مثبت پیشوند را (مانند بیتا، ناهید، انوشه و ...) بتوان در نظر گرفت. در حقیقت بدین وسیله نام‌ها به سه گونه یا نژاد تقسیم‌بندی شده‌اند که (۱) نشانگر بار مثبت، (۰) نشانگر بار خنثی، و (۱-) نشانگر بار منفی بخش معنایی نخست نام بوده و در این پژوهش آمیزش گونهٔ سوم با دو گونهٔ دیگر مجاز نبوده است.

از آن‌جاکه در این پژوهش نام‌های ایرانی باید از نظر آوایی بررسی شوند، بر پایهٔ روش به‌کار رفته در عروض قافیه اشعار فارسی[۳۱]، هجاهای تشکیل‌دهندهٔ هر یک از نام‌ها نیز تجزیه شدند. جدول ۷ نمادها و چگونگی این روش را نشان می‌دهد. همچنین در این پایگاه برای کوته‌نویسی و سادگی کار، نمادهای x، y و z (بزرگ و کوچک) منظور شده و هجاهای بلند و کشیده نیز برابر در نظر گرفته شده‌اند. در این جدول c نماد حروف همخوان، v نماد واکه‌های کوتاه و V نماد واکه‌های بلند است.

البته دو هجای cvccc و cVccc (مانند سانتر و لوستر) به‌دلیل کمبود کاربرد در فارسی، نادیده گرفته شده است. بر پایهٔ این هجابندی، فیلدهای جدیدی به پایگاه داده‌ای نام‌ها افزوده شد که در جدول ۸ نمایش داده شده است.

نکتهٔ مهم آن‌که در این پژوهش، همان‌گونه که گفته شد، تفکیک معنایی واژگان تشکیل‌دهندهٔ نام‌ها، برای تعیین محل گسست و ترکیب ژن‌ها در الگوریتم ژنتیک انجام شده، اما تفکیک هجایی نام‌ها برای استفاده در شبکهٔ عصبی به‌منظور تعیین میزان استفادهٔ هر نوع هجا و در نتیجه شبیه‌سازی زیبایی‌شناسی هجاها در آن، انجام پذیرفته است.

در ادامه، برای تعیین بردار برازندگی هر نمونه در الگوریتم ژنتیک و نیز خروجی هر نمونه در شبکه عصبی، فیلدهای میانگین درجه زیبایی هر نمونه از دیدگاه آوایی و معنایی به‌همراه تعداد ثبت‌نام‌ها در سازمان ثبت احوال کشور[۳۵]، به پایگاه نام‌ها افزوده شدند. برای آن‌که مقایسه این فیلدها امکان‌پذیر باشد و نیز برای آن‌که ورودی‌های بردار هدف y شبکه عصبی مقادیری نرمال داشته باشند، این میانگین‌ها به‌کمک رابطهٔ زیر به مقادیری در بازه (۱+ و ۱-) نگاشت شدند.

$$y = \frac{x - mid(\bar{X})}{(max(\bar{X}) - min(\bar{X}))/2}$$

رابطه (۱)

در رابطهٔ (۱)، متغیر x مقدار فیلد میانگین امتیازات آوایی، معنایی یا ثبت‌احوال و بردار X نشانگر کل مقادیر این فیلدها و y خروجی نگاشت شده هر یک از آنهاست.

سرانجام برای بررسی و مقایسه اختلاف‌نظر جامعه آماری این پژوهش (که پرسش‌نامهٔ نام‌ها میان آنان توزیع شده) با میزان درخواست‌های ثبت‌شده در پایگاه وب سازمان ثبت احوال (که به‌نوعی استقبال عمومی از آن نام‌ها را نشان می‌دهد)، میانگین این عامل‌ها از رابطه زیر محاسبه شد:

$$t = |(v + m)/2|$$

رابطه (۲)

که در این رابطه v میانگین امتیاز آوایی نگاشت‌شده، m میانگین امتیاز معنایی نگاشت‌شده، و t میانگین کل آنهاست. جدول ۹ نمونه‌هایی از این مقادیر را برای چند نام نمایش می‌دهد.

البته فیلد تفاضل، با میانگین ۰/۷۹ و انحراف معیار ۰/۴۵ همبستگی خاصی را میان امتیازات آوایی و معنایی تخصیص‌یافته از سوی جامعه آماری در دانشگاه پرند، و آمار ثبت‌شده در سازمان ثبت احوال کشور نشان نمی‌دهد.

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

**جدول ۶) پایگاه نام‌ها، به‌همراه آوانگاری و تفکیک آوایی و معنایی آنها.**

| ردیف | نام | معنا | آوانگاری | هجای۱ | هجای۲ | هجای۳ | معنای۱ | معنای۲ | معنای۳ | بار |
|---|---|---|---|---|---|---|---|---|---|---|
| ۱ | اردشیر | پادشاهی پاک / کشور مقدس | &ardeSir | &ar | de | Sir | &ard | Sir | | 1 |
| ۲ | اردوان | نگهبان راستی / پاسدار درستکاران | &ardavAn | &ar | da | vAn | &ard | vAn | | 1 |
| ۳ | ارژنگ | جادویی / زینت / نقش و نگار | &rZang | &r | Zang | | &ar | Zang | | 0 |
| ۴ | البرز | از همه سو بلند / سراسر گسترده | &alborz | &al | borz | | &al | borz | | 0 |
| ۵ | الوند | دارای تندی و تیزی / تندمند | &alvand | &al | vand | | &al | vand | | 0 |
| ۶ | اندیشه | از همه سو نگریستن / فکر | &andiSe | &an | di | Se | &an | diS | e | 0 |
| ۷ | انوشه | بی‌مرگ / جاودان | &anuSe | &a | nu | Se | &a | nuS | e | -1 |
| ۸ | ایزدیار | آنکه خدا بااوست / دوست خدا | &izadyAr | &i | zad | yAr | &izad | yAr | | 1 |
| ۹ | آرشیدا | فروغ آریایی / خورشید ایران | &ArSidA | &Ar | Si | dA | &Ar | Sid | A | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

**جدول ۷) انواع هجا در زبان فارسی [۳۱].**

| ردیف | نماد هجا | نماد در | نمونه |
|---|---|---|---|
| ۱ | cv | x | نه، به، تو |
| ۲ | cvc | y | تر، دل، گل |
| ۳ | cvcc | z | سرد، مهر، دزد |
| ۴ | cV | X | با، سی، رو |
| ۵ | cVc | Y | سال، میز، دود |
| ۶ | cVcc | Z | پارس، سیرک، دوخت |

**جدول ۸) آوانگاری عروضی نام‌ها**

| ردیف | نام | آوا | آوای هجای۱ | آوای هجای۲ | آوای هجای۳ | نماد هجای۱ | نماد هجای۲ | نماد هجای۳ |
|---|---|---|---|---|---|---|---|---|
| ۱ | اردشیر | cvccvcVc | cvc | cv | cVc | x | y | Y |
| ۲ | اردوان | cvccvcVc | cvc | cv | cVc | x | y | Y |
| ۳ | ارژنگ | cvccvcc | cvc | cvcc | | z | y | |
| ۴ | البرز | cvccvcc | cvc | cvcc | | z | y | |
| ۵ | الوند | cvccvcc | cvc | cvcc | | z | y | |
| ۶ | اندیشه | cvccVcv | cvc | cV | cv | X | y | x |
| ۷ | انوشه | cvcVcv | cv | cV | cv | X | x | x |
| ۸ | ایزدیار | cVcvccVc | cV | cvc | cVc | X | y | Y |
| ۹ | آرشیدا | cVccVcV | cVc | cV | cV | Y | X | X |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

**جدول ۹) امتیازات آوایی و معنایی و فراوانی نام‌ها در پایگاه وب سازمان ثبت احوال کشور و مقادیر نگاشت شدهٔ معادل آنها**

| نام | میانگین امتیازات آوایی | میانگین امتیازات معنایی | فراوانی در پایگاه وب سازمان ثبت | میانگین امتیازات (نگاشت شده) آوایی | میانگین امتیازات (نگاشت شده) معنایی | فراوانی در سازمان ثبت (نگاشت‌شده) | تفاضل |
|---|---|---|---|---|---|---|---|
| اردشیر | ۵,۳۴ | ۳,۹ | ۱۷۴ ۲۴۱۳۳ | -۰,۳۶۶۷ | ۰,۱۱۳۳ | -۰,۷۷۹۰ | ۰,۶۵۲۴ |
| اردوان | ۵,۲۲ | ۴,۱۶ | ۲۹۶۸ | -۰,۲۸۰۰ | ۰,۰۷۳۳ | -۰,۹۷۳۱ | ۰,۸۶۹۸ |
| ارژنگ | ۴,۲۸ | ۳,۳۲ | ۱۸۸۱ | -۰,۵۶۰۰ | -۰,۲۴۰۰ | -۰,۹۸۳۰ | ۰,۵۸۳۰ |
| البرز | ۵,۲۸ | ۴,۲۲ | ۲۹۱۷ | -۰,۲۶۰۰ | ۰,۰۹۳۳ | -۰,۹۷۳۶ | ۰,۸۹۰۲ |
| الوند | ۴,۲ | ۳,۸ | ۱۹۶ | -۰,۴۰۰۰ | -۰,۲۶۶۷ | -۰,۹۹۸۵ | ۰,۶۶۵۱ |
| اندیشه | ۵,۶۶ | ۴,۰۴ | ۶۷۰ | -۰,۳۲۰۰ | ۰,۲۲۰۰ | -۰,۹۹۴۱ | ۰,۹۴۴۱ |
| انوشه | ۵,۹۸ | ۳,۶۴ | ۹۴۰ | -۰,۴۵۳۳ | ۰,۳۲۶۷ | -۰,۹۹۱۶ | ۰,۹۲۸۳ |
| ایزدیار | ۶,۲۸ | ۳,۵۶ | ۷۰ | -۰,۴۸۰۰ | ۰,۴۲۶۷ | -۰,۹۹۹۶ | ۰,۹۷۲۹ |
| آرشیدا | ۷ | ۵,۵ | ۳۳۱ | ۰,۱۶۶۷ | ۰,۶۶۶۷ | -۰,۹۹۷۲ | ۱,۴۱۳۹ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

### ۴-۴- شبکه عصبی به‌عنوان تابع برازندگی

برای دست‌یابی به بهترین نتیجه، انواع شبکه‌های عصبی موجود در جعبه‌ابزار شبکه عصبی متلب نگارش ۷/۸ [۱۳] با مجموعه‌های ورودی‌های گوناگون آزموده شدند که سرانجام با توجه به این تجربیات و نیز توانمندی‌ها و ویژگی‌های انواع شبکه‌های عصبی[۲۹]، شبکهٔ عصبی پیش‌خور پس‌انتشار[۱۴] با ۲ لایه پنهان، هر لایه با نرون‌هایی به‌تعداد ۲ برابر بردار ورودی، و یک نرون خروجی دوقطبی، و تابع آموزش از پس‌انتشار لونبرگ-مارکوارت[۱۵](TRAINLM) و تابع فعالیت تانژانت هایپربولیک سیگموید (TANSIG) در همه لایه‌ها (شکل ۵) به بهترین نتیجه رسید و برگزیده شد. این شبکه در حقیقت یک شبکهٔ پرسپترون چندلایه[۱۶] است که به‌طور نظری ثابت در صورت داشتن دست‌کم یک لایهٔ میانی و تابع تبدیل سیگموید و یک تابع تبدیل خطی در لایهٔ خروجی، می‌تواند همهٔ توابع دلخواه را تقریب بزند؛ مشروط بر آن‌که دارای نرون‌های کافی در لایه میانی باشد. البته محدودیت‌های شبکه‌های عصبی پرسپترون چند لایه دو مورد است: نخست این‌که تابع‌های مورد تقریب، باید پیوسته باشند و دوم، شمار نرون‌های لایهٔ پنهان را نمی‌توان نامحدود برگزید[۳۶و۳۷].

#### ۴-۴-۱- بردارهای ورودی و هدف

برای دست‌یابی به بهترین نتیجه، ورودی‌ها به‌صورت‌های گوناگون (به‌کمک برنامهٔ مبدل در محیط ویژوال بیسیک) به کدهای ورودی شبکهٔ عصبی تبدیل و آزموده شد که برخی از موثرترین آنها در ادامه معرفی می‌شود.

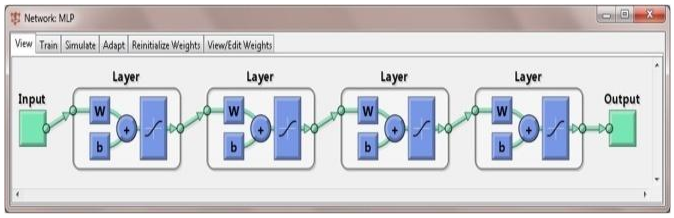
#### ۴-۴-۲- بردار خروجی

همان‌گونه که پیشتر توضیح داده شد، میانگین نظرات جامعهٔ آماری در مورد زیبایی نام‌ها، به‌عنوان مقادیر بردار هدف (خروجی) شبکهٔ عصبی در نظر گرفته شدند که البته برای افزایش دقت و درستی شبکه، این مقادیر بر پایهٔ رابطهٔ (۱) در بازه (۱+ و ۱-) نگاشت شدند.

#### ۴-۴-۳- ورودی به‌صورت نوع هجا

برای تشخیص این‌که کدام هجاهای زبان فارسی از دیدگاه جامعهٔ آماری زیباتر بوده است، آن‌گونه که در جدول ۶ و ۸ نمایش داده شده، هر نوع هجا بنابر قاعدهٔ هجاهای فارسی، منطبق با علم عروض ادبی[۳۱]، تفکیک شد و به انواع آن که در جدول ۷ نگاشت شده، کدی تخصیص یافت که کد حاصله در جدول ۱۰ موجود است.

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026




**شکل ۵) نمای کلی شبکهٔ عصبی به‌کار رفته در پروژه**

**جدول ۱۰) کد ورودی معادل هر یک از هجاهای فارسی به شبکه**

| ردیف | نماد هجا | نماد در | کد (دوقطبی) |
|---|---|---|---|
| ۱ | cv | x | -1 -1 -1 -1 -1 |
| ۲ | cvc | y | -1 -1 -1 -1 +1 |
| ۳ | cvcc | z | -1 -1 -1 +1 -1 |
| ۴ | cV | X | -1 -1 +1 -1 -1 |
| ۵ | cVc | Y | -1 +1 -1 -1 -1 |
| ۶ | cVcc | Z | +1 -1 -1 -1 -1 |

**جدول ۱۱) مبنای واکه‌های فارسی**

| ردیف | آوا | نماد | کد (دوقطبی) |
|---|---|---|---|
| ۱ | اَ | a | -1 -1 -1 -1 -1 +1 |
| ۲ | اِ | e | -1 -1 -1 -1 +1 -1 |
| ۳ | اُ | o | -1 -1 -1 +1 -1 -1 |
| ۴ | آ | A | -1 -1 +1 -1 -1 -1 |
| ۵ | ایـ | i | -1 +1 -1 -1 -1 -1 |
| ۶ | او | u | +1 -1 -1 -1 -1 -1 |

این شبکه نه‌تنها ارزش یک نوع هجا، که ارزش توالی آنها را نیز نشان می‌دهد. برای نمونه، شبکه‌ای با این ورودی‌ها می‌تواند بین دو ترکیب مانند xX و Xx (یا cvcV و cVcv)، مانند ساره و رسا، تمایز قایل شود، اما نمی‌تواند نوع واکه‌ها و همخوان‌های کوتاه و بلند آن را تشخیص دهد. ورودی این شبکه برای حداکثر ۴ هجا (۲۴ گره ورودی) تنظیم شد و گره‌های هجاهای خالی با مقدار (1-) پر شدند.

پس از آموزش و آزمایش متوالی شبکهٔ عصبی با چنین ورودی‌هایی که نمونه‌های آموزشی و آزمایش آن به‌طور متوالی و تصادفی از ۲۵۲ نام پایگاه اولیه انتخاب می‌شد، میانگین اختلاف خروجی شبکه با خروجی‌های واقعی (هدف) به حدود ۰/۷۲ در بازه (۱+ و ۱-) یا به‌عبارت بهتر خطای ۳۶٪ (دقت ۶۴٪) رسید که با توجه به کمینه بودن ورودی‌ها نتیجه چندان بدی هم به‌شمار نمی‌رود.

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

**۴-۴-۴- ورودی برپایهٔ نحوهٔ تولید حرف در هجا**

در روشی دیگر، برای افزایش دقت شبکه، ورودی‌ها بر پایهٔ نحوهٔ تولید همخوان‌ها منطبق با جدول ۳ و واکه‌ها منطبق با جدول ۵ تنظیم شد که کد معادل آنها در جدول ۱۱ و ۱۲ نمایش داده شده است.

گفتنی است که کدهای بر پایهٔ شانزده مندرج در جدول ۸ تنها برای سادگی نمایش داده شده و در عمل به‌صورت دودویی-دوقطبی ۱۴ بیتی (رقومی) به شبکه عصبی وارد گشته است. این کدها، برای افزایش کارایی شبکهٔ عصبی، به‌گونه‌ای تنظیم شده که هر سطر یا ستون در جدول ۳ یک بیت (رقم) با خانه همسایه‌اش اختلاف داشته باشد. جدول ۱۳ مبنای این کدگذاری را نشان می‌دهد که مرکز آن در جدول ۳، فیلد روان- لثوی و کامی قرار گرفته است. دو رقم کم‌ارزش در کدهای جدول ۱۲ نشانگر ردیف و ۲ رقم پرارزش نشانگر ستون هر فیلد است.

**جدول ۱۲) کد معادل هر یک از نحوه/محل‌های تولید واکه‌های جدول۳**

| ردیف | نحوه - محل تولید | کد (برپایه ۱۶) |
|---|---|---|
| ۱ | انفجاری - دولبی | 0F0F |
| ۲ | انفجاری - دندانی | 0C0F |
| ۳ | انفجاری - کامی | 100F |
| ۴ | انفجاری - ملازی | 300F |
| ۵ | انفجاری - چاکنایی | 700F |
| ۶ | سایشی - لب و دندانی | 0E0E |
| ۷ | سایشی - لثوی | 080E |
| ۸ | سایشی - لثوی و کامی | 000E |
| ۹ | سایشی - ملازی | 300E |
| ۱۰ | سایشی - چاکنایی | 700E |
| ۱۱ | انفجاری/سایشی - لثوی و کامی | 000C |
| ۱۲ | لرزشی - دندانی | 0C08 |
| ۱۳ | روان - دندانی | 0C00 |
| ۱۴ | روان - کامی | 1000 |
| ۱۵ | غنه‌ای - دولبی | 0F10 |
| ۱۶ | غنه‌ای - دندانی | 0C10 |

**جدول ۱۳) مبنای کدگذاری جدول ۸**

| ردیف | کد باینری | کد (برپایه ۱۶) |
|---|---|---|
| ۱ | 0001111 | 0F |
| ۲ | 0001110 | 0E |
| ۳ | 0001100 | 0C |
| ۴ | 0001000 | 08 |
| ۵ | 0000000 | 00 |
| ۶ | 0010000 | 10 |
| ۷ | 0110000 | 30 |
| ۸ | 1110000 | 70 |

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

پس از آموزش و آزمایش متوالی شبکهٔ عصبی با چنین ورودی‌هایی، میانگین اختلاف خروجی شبکه با خروجی‌های واقعی (هدف) به حدود ۰/۲۵ در بازه (۱+ و ۱-)، یعنی خطای ۱۲/۵٪ (دقت ۸۷/۵٪) رسید که دقتی مطلوب به‌شمار می‌رود.

### ۴-۴-۵- ورودی به‌صورت واج

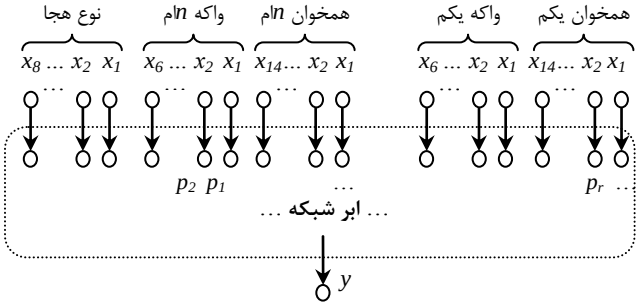
ورود نام‌ها بر پایهٔ نحوهٔ تولید حروف، دقت شبکه را بسیار افزایش داد، اما این ورودی هم‌چنان نمی‌تواند تفاوت برخی از حروف را بازشناسی کند. برای نمونه، کدهای دو نام دارا و تارا هیچ تفاوتی با یکدیگر ندارد؛ هرچند که ممکن است زیبایی آوایی آنها از دید افراد متفاوت باشد. از این‌رو به‌نظر می‌رسید که تهیهٔ ورودی شبکه به‌تفکیک واج‌های هر نام، دقت کار را افزایش دهد.

در این روش نیز به هر یک از ۲۹ واج فارسی (۲۳ همخوان و ۶ واکه) مندرج در جدول ۵، کدی یکتا اختصاص یافت و آموزش و آزمایش ترکیبات مختلف شبکهٔ عصبی انجام پذیرفت. اما در بهترین ترکیب شبکه، میانگین خطای ۱۴٪ (دقت ۸۶٪) به‌دست آمد که هرچند دقتی مطلوب به‌شمار می‌آید، اما از ورودی بر پایهٔ نحوهٔ تولید حروف کمتر است.

عدم افزایش دقت شبکه با چنین ورودی‌هایی ممکن است به‌دلیل کمبود نمونه‌ها، ترکیب نادرست شبکه یا کدهای ورودی، بالا بودن بار شبکه (به ازای یک نام ۷ واجی، ۲۰۳ گره ورودی) یا دلایل دیگر بوده باشد که به‌هر ترتیب امکان برطرف ساختن این مشکلات فراهم نشد.

### ۴-۴-۶- ورودی ترکیبی

از میان آزمایش‌های گوناگون دیگر، دو روش ورودی به‌صورت نوع هجا و بر پایهٔ نحوهٔ تولید حروف با یکدیگر ترکیب شد و در نتیجه دقت شبکه به حدود ۸۹٪ رسید. این شبکه که طرح کلی آن در شکل ۶ نمایش داده شده، به‌عنوان راه‌حل نهایی برگزیده شد.




**شکل ۶) تولید نام‌های جدید به‌کمک الگوریتم ژنتیکپ**

**جدول ۱۴) نسبت واکه‌ها با یکدیگر برای محاسبهٔ احتمال دگرگونی آنها با استفاده از رابطهٔ ۳**

| واکه‌ها | اَ (a) | آ (A) | اِ (e) | ایـ (i) | اُ (o) | او (u) | حذف واکهٔ آغاز یا پایان |
|---|---|---|---|---|---|---|---|
| اَ (a) | - | x | y | z | z | z | x |
| آ (A) | x | - | z | z | z | z | x |
| اِ (e) | y | z | - | x | z | z | x |
| ایـ (i) | z | z | x | - | z | z | x |
| اُ (o) | z | z | z | z | - | x | y |
| او (u) | z | z | z | z | x | - | y |
| ـه (a/e) | x | x | - | x | z | z | x |

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

### ۴-۵- الگوریتم ژنتیک

الگوریتم ژنتیک به‌کار رفته در این پروژه نسبتا ساده بوده است. بدین ترتیب که بر نام‌هایی که از ۲ تا ۴ تکواژ تشکیل شده‌اند، فاز بازتولید (تولید مثل) صفر تا ۲ بار (بر اساس برازندگی) اعمال شده و بر هر یک از تکواژها نیز صفر تا ۲ بار جهش (با مقدار احتمال قابل تنظیم) رخ داده است. در زیر این روند برای نام‌های تولیدی سایمان، سورین و بینوش نمایش داده شده است:

$$\left\{\begin{matrix}\text{sAyna} = \text{sAy} + \text{nA}\\ \text{human} = \text{hu} + \text{man}\end{matrix}\right\} \xrightarrow{\text{cross over}} \{\text{sAy} + \text{man}\} \xrightarrow{\text{mutation}} \{\text{sAy} + \text{mAn}\} \xrightarrow{\text{output}} \text{sAymAn}$$

$$\{\text{surenA} = \text{suren} + \text{A}\} \xrightarrow{\text{mutation}} \{\text{surin} + \text{A}\} \xrightarrow{\text{mutation}} \{\text{surin} + \emptyset\} \xrightarrow{\text{output}} \text{surin}$$

$$\left\{\begin{matrix}\text{bitA} = \text{bi} + \text{tA}\\ \text{anuSe} = \text{a} + \text{nuS} + \text{e}\end{matrix}\right\} \xrightarrow{\text{cross over}} \{\text{bi} + \text{nuS} + \text{e}\} \xrightarrow{\text{mutation}} \{\text{bi} + \text{nush} + \emptyset\} \xrightarrow{\text{output}} \text{binuS}$$

از آنجا که در طول تاریخ، تعداد دگرگونی هر یک از واکه‌ها و همخوان‌ها به‌یکدیگر، به‌طور مساوی رخ نداده است، احتمال رخداد جهش و نیز احتمال رخداد دگرگونی یک واکه به واکهٔ دیگر (مانند a به e)، بر پایهٔ جدول ۲ به‌فرم جدول ۱۴ به‌وسیلهٔ رابطهٔ ۳ محاسبه می‌شود. اما می‌توان برای به‌دست آورد مقادیر حقیقی‌تر، پردازشی گسترده‌تر بر سیر تطور واژگان زبان فارسی انجام داد؛ که البته از موضوع این پژوهش بیرون بوده است. در آزمون این طرح، صدها نام تازه پدید آمد که هر یک از آنها برای دریافت مقدار تابع برازندگی به شبکهٔ عصبی وارد و به پایگاه داده (جمعیت) افزوده می‌شد.

### ۴-۶- نظرسنجی نام‌های جدید از جامعه آماری

پس از تولید انبوه نام‌های جدید، ۴۰ نام که بالاترین رتبه (برازندگی) را توسط شبکهٔ عصبی دریافت کرده بودند، در قالب پرسش‌نامه‌ای مانند پرسش‌نامهٔ نمونه‌های اولیه (مندرج در پیوست ۱) به همان جامعه آماری ارایه و نظرات آنان دربارهٔ نام‌های جدید گردآوری و از دیدگاه آماری پردازش شد.

### ۴-۷- تحلیل و ارزیابی

از میان نام‌های برگزیده، نزدیک به ۳۵٪ از نام‌های تولید شده، از نام‌های شناخته شده توسط سازمان ثبت احوال کشور و حتی ۳٪ از نام‌ها در جامعهٔ اولیه موجود بوده‌اند که این امر می‌تواند نشان از درستی روش باشد. برای ۴۰ نام برتر، وجود نام در پایگاه وب سازمان ثبت احوال کشور به‌همراه تعداد تکرار آن و نیز میانگین امتیازات آوایی و معنایی جامعهٔ آماری به هر نام جدول ۱۴ درج شده است.

میانگین نظرات جامعهٔ آماری در مورد هر نام تازه، نسبت به خروجی شبکهٔ عصبی (برازندگی) همان نام، که در بازهٔ (۱ و ۰) نگاشت شده، اختلافی با میانگین ۰/۲۳ و انحراف معیار ۰/۱۲ را نشان می‌دهد که نشانگر همبستگی مناسبی بین این دو عامل است.

همچنین از این ۴۰ نام، نامی که بالاترین امتیاز را در میان اعضای جامعه دریافت کرده (آرتینا)، رتبهٔ ۵ را در میان ۲۵۲ نام پایگاه آغازین که توسط همان افراد ارزش‌دهی شده، کسب کرده که جایگاهی بسیار درخور توجه است. چهلمین نام جدید (جمداد) نیز رتبه ۲۰۱ را به خود اختصاص داده است. رتبهٔ نام‌های جدید نیز در پایگاه نام‌های آغازین، در جدول ۱۵ نمایش داده شده است.

## ۵- جمع‌بندی و کارهای آینده

در این پژوهش، با شناخت روند دگرگونی و تکامل نام‌ها و واژگان ایرانی در طول تاریخ، و شبیه‌سازی هوشمند این روند، نام‌های جدید معنادار و به‌نسبت زیبایی به‌کمک الگوریتم ژنتیک تولید شد. زیبایی آوایی و معنایی نام‌هایی که نقش جامعهٔ آغازین الگوریتم ژنتیک را داشتند، توسط یک جامعهٔ آماری که بیشترین افرادش در دانشگاه آزاد اسلامی واحد پرند اشتغال داشته‌اند، تعیین شد. این نام‌ها یک شبکه عصبی مصنوعی پیشخور پس‌انتشار را که توانست تا ۸۹٪ نظرات جامعهٔ آماری را شبیه‌سازی کند، آموزش

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

دادند تا آن شبکه نقش تابع برازندگی را برای به‌دست آوردن میزان زیبایی آوایی نام‌ها در الگوریتم ژنتیک، ایفا کند. جامعهٔ آغازین در الگوریتم ژنتیک را مجموعه‌ای از نام‌های ایرانی مرکب از ۲ یا ۳ بن‌واژه تشکیل می‌دادند که محل گسست/پیوست ژن‌ها و احتمال و چگونگی رخداد جهش ژنتیکی بر پایهٔ قواعد زبان‌شناسی زبان‌های ایرانی تدوین شد. بهترین نام‌های تولید شده بدین روش، رتبه‌های ۵، ۷ و ۸ را در میان ۲۵۲ نام پایگاه نام‌های آغازین به‌دست آورده‌اند.

پیشنهاد برای بهبود و افزایش کارایی پژوهش:

۱- نام‌های جامعهٔ آغازین افزایش یابد.

۲- پژوهش بر دیگر واژگان فارسی (نه‌فقط نام‌ها) انجام شود.

۳- دیگر روش‌های هوش مصنوعی و دیگر انواع شبکهٔ عصبی آزموده شود.

۴- پژوهشی جامع در زمینهٔ چگونگی دگرگونی تاریخی واکه‌ها و همخوان‌ها به یکدیگر انجام پذیرد.

**جدول ۱۵) ۴۰ نام برتر تولید شده توسط الگوریتم ژنتیک و برگزیده توسط شبکه عصبی**

| ردیف | نام | معنا | آوانگاری | فراوانی در سایت ثبت | امتیاز معنایی | امتیاز آوایی | به در پایگاه نام |
|---|---|---|---|---|---|---|---|
| ۱ | **اردکام** | دارای زندگی و سرنوشت مقدس | &ardakAm | ۴ | ۶/۷۸ | ۴/۹۷ | ۷۰ |
| ۲ | **آتراب** | آب و آتش | AtrAb | - | ۵/۲۵ | ۳/۷۵ | ۱۸۰ |
| ۳ | **آتراس** | دارای اسب آتشین | AtrAs | - | ۴/۷۲ | ۴/۰۶ | ۱۸۶ |
| ۴ | **آترین** | آتشین | Atrin | ۳۱۸ | ۶/۱۹ | ۶ | ۶۱ |
| ۵ | **آرتینا** | پاک و مقدس | &ArtinA | ۲۹۶ | ۷/۷۸ | ۶/۸۸ | ۵ |
| ۶ | **آردیشا** | دارای بینش و نگرش آریایی | ArdiSA | - | ۵/۷۲ | ۴/۸۸ | ۱۲۰ |
| ۷ | **آرشید** | فروغ آریایی / ایرانی درخشان | &ArSid | - | ۷/۳۴ | ۵/۹۱ | ۳۱ |
| ۸ | **آرمن** | دارای اندیشه آریایی | &Arman | - | ۶/۱۳ | ۵/۶۶ | ۶۷ |
| ۹ | **آرنگ** | ایرانی اصیل / دارای روی سپید | Arang | ۲۶ | ۶/۷۲ | ۵/۲۵ | ۶۳ |
| ۱۰ | **آروان** | نگهبان ایران / پاسدار ایرانیان اصیل | &ArvAn | ۳۳ | ۶/۶۶ | ۵/۹۷ | ۵۰ |
| ۱۱ | **باروش** | لطیف مانند باران | bAroS | - | ۶/۷۸ | ۴/۶۹ | ۸۱ |
| ۱۲ | **بردین** | دارای دین و عقیده والا | bardin | - | ۵/۹۱ | ۵/۲۸ | ۹۲ |
| ۱۳ | **بهزیدا** | نقش و نگار ظریف و نیکو | behzidA | - | ۵/۶۹ | ۳/۹۷ | ۱۵۱ |
| ۱۴ | **بهفام** | خوش آب و رنگ / زیبا | behfAm | ۷ | ۶/۱۹ | ۵/۰۳ | ۹۰ |
| ۱۵ | **بینوش** | جاودان / بی‌مرگ | binuS | - | ۶/۸۱ | ۴/۷۵ | ۷۵ |
| ۱۶ | **پارسینا** | فرد پارسا و درستکار / ایرانی اصیل | pArsinA | ۲۹ | ۷/۱۶ | ۶/۱۹ | ۲۸ |
| ۱۷ | **تارین** | درخشان همچو ستاره | tArin | - | ۶/۳۱ | ۵/۷۸ | ۶۱ |
| ۱۸ | **جمداد** | داده فروغ / عدالت روشن | jamdAd | - | ۴/۸۴ | ۳/۳۸ | ۲۰۱ |
| ۱۹ | **دلسین** | شجاع / دارای دلی قرص و درخشان | delsin | - | ۵/۹۴ | ۴/۳۱ | ۱۲۶ |
| ۲۰ | **رخشید** | دارای چهره درخشان | roxSid | ۲۳ | ۶/۰۶ | ۴/۳۱ | ۱۲۱ |
| ۲۱ | **رستاهید** | رهاننده از ناپاکی و پلیدی | rastAhid | - | ۶/۵ | ۳/۷۸ | ۱۲۵ |
| ۲۲ | **روژوان** | نگهبان روز / خورشید | roZvAn | - | ۶/۰۹ | ۵/۳۱ | ۸۵ |
| ۲۳ | **سایمان** | اندیشه درخشان | sAymAn | - | ۶/۷۸ | ۵/۵ | ۶۰ |
| ۲۴ | **سورین** | نیرومند / توانا | surin | ۲۱۵ | ۶/۹۷ | ۵/۹۷ | ۴۰ |

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ۲۵ | **شران** | شیرین | SerAn | - | ۵/۹۴ | ۴/۲۵ | ۱۲۹ |
| ۲۶ | **شهسا** | مانند پادشاهان | SahsA | - | ۶/۴۴ | ۵/۵۹ | ۶۲ |
| ۲۷ | **شهفر** | دارای شکوه پادشاهان | Sahfar | - | ۵/۵۹ | ۴/۳۱ | ۱۴۰ |
| ۲۸ | **شهکامه** | دارای بخت شاهان / نیک بخت | SahkAme | - | ۶/۱۹ | ۴/۸۴ | ۱۰۳ |
| ۲۹ | **شیرا** | شیرین | SirA | - | ۵/۵۹ | ۴/۴۱ | ۱۳۷ |
| ۳۰ | **فرلان** | بلند مرتبه/ دارای جایگاه شکوهمند | farlAn | - | ۷/۱۹ | ۵/۰۶ | ۶۰ |
| ۳۱ | **کوران** | تابش خورشید / آفتاب | kurAn | - | ۵/۶۳ | ۳/۲۸ | ۱۸۶ |
| ۳۲ | **کیارام** | دارای رفتار باوقار همچون پادشاهان | keyArAm | ۷ | ۶/۹۷ | ۶/۳۱ | ۳۰ |
| ۳۳ | **گلتین** | باغبان / دارنده گلهای بسیار | goltin | - | ۵/۵ | ۵/۰۶ | ۱۲۰ |
| ۳۴ | **مهبد** | نگهبان ماه / سرور و والا | mahbod | ۱۴۷۶ | ۷/۱۳ | ۷/۴۴ | ۷ |
| ۳۵ | **مهراد** | والاترین جوانمرد | mehrAd | ۲۷۷۹ | ۷/۳۴ | ۷/۱۹ | ۸ |
| ۳۶ | **مه‌فروز** | درخشان همچو ماه | mahforuz | ۱۳ | ۶/۲۲ | ۴/۷۵ | ۱۰۵ |
| ۳۷ | **نوش‌آرا** | فراهم‌آورنده آسایش و شادی | nuS&ArA | - | ۶/۴۱ | ۵/۴۷ | ۶۵ |
| ۳۸ | **هوراد** | جوانمرد نیک | hurAd | ۷۴ | ۶/۵۹ | ۵/۴۴ | ۶۲ |
| ۳۹ | **هومین** | خوش معنا / دارای معنویت خوب | humin | - | ۶/۳۸ | ۴/۵ | ۱۰۸ |
| ۴۰ | **ویستم** | دارای دانش و توانایی | vistam | - | ۶/۲۵ | ۳/۵ | ۱۴۷ |

**منابع**


۱.مهدی حمیدا، فرهنگ نام‌ها، کلک آزادگان، چاپ سوم، ۱۳۸۶.

۲.جهانگیر اوشیدری، دانشنامه مزدیسنا، نشر مرکز، چاپ سوم، ۱۳۸۳.

۳.حسین شهیدی مازندرانی، فرهنگ شاهنامه (نام کسان و جای‌ها)، نشر بلخ، چاپ نخست، ۱۳۷۷.

۴.ا. ش. شاهمیری، ر. دژکام و س. شیری، شناسایی اشعار شاهنامه فردوسی به‌کمک شبکه عصبی مصنوعی، نشریهٔ انجمن کامپیوتر ایران (JCSE 2007 )، مجلد ۴، شماره ۳ (الف) پاییز ۱۳۸۵، ص ۱۷-۲۶.

۵.ا. ش. شاهمیری و م. ر. مطش بروجردی، تعیین شاعر به‌کمک روش‌های یادگیری ماشین، سومین کنفرانس فناوری اطلاعات و دانش (IKT 2007) ، مشهد، ۱۳۸۵.

۶.ا. ش. شاهمیری، م. ر. صفابخش و ر. دژکام، تصحیح خودکار غلط‌های تایپی فارسی به‌کمک شبکه عصبی مصنوعی ترکیبی، نشریهٔ علمی پژوهشی انجمن مهندسین برق و الکترونیک ایران، سال پنجم، شماره یکم، بهار و تابستان ۱۳۸۷، ص ۱۶-۲۹.

۷.ا. ش. شاهمیری، م. ر. صفابخش و ر. دژکام، تعیین ریشه زبانی واژگان فارسی و عربی به‌کمک شبکه عصبی مصنوعی، دوازدهمین کنفرانس بین‌المللی انجمن کامپیوتر ایران (CSICC 2007)، تهران، ۱۳۸۵.

۸.ع. ح. اصفهانی، س. راحتی قوچانی و ن. جهانگیری، استخراج ویژگی برای یک سیستم شناسایی و طبقه‌بندی اسامی فارسی، پانزدهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران، تهران، ۱۳۸۸.

۹.نادیه سجادی و مهرنوش شمس‌فرد، تبدیل متن محاوره‌ای فارسی به رسمی به‌کمک N-gramها، شانزدهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران، تهران، ۱۳۸۹.

۱۰.ن. ریاحی، ف. غزالی و م. ع. غزالی، سامانه خلاصه‌ساز متن‌های فارسی با رویکرد ترکیبی شبکه عصبی RBF و الگوریتم ژنتیک، هجدهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران، تهران، ۱۳۹۱.

۱۱.ا. بابادی، پ. غیاث‌نژاد عمران و غ. قاسم‌ثانی، استفاده از

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

یادگیری ماشین در ریشه‌یابی کلمات فارسی، هجدهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران، تهران، ۱۳۹۱.
۱۲.م. ص. زاهدی، ا. بزرگی و ک. فاتحی، بررسی ریشه‌یاب‌های واژگان زبان فارسی و تاثیر آنها در کارایی سیستم‌های بازیابی اطلاعات متنی، نخستین کنفرانس بین‌المللی پردازش زبان و خط فارسی، سمنان، ۱۳۹۱.
۱۳.مهربان‌گشتاسپ پورپارسی، گنجینهٔ نام‌های ایرانی، انتشارات فروهر، چا دوم، ۱۳۶۲.
۱۴.ع. زاهدی‌نیا و ک. کاظمی، فرهنگ نام ایرانیان، انتشارات ترانه، چاپ چهارم، ۱۳۸۶.
۱۵.سودابه مبشر، ایرانیان فرزندان خود را چه می‌نامند؟ (فرهنگ جامع نام‌های ایرانی)، نشر پارسه، چاپ پنجم، ۱۳۸۹.
۱۶.پری زنگنه، آوای نام‌ها از ایران‌زمین، کتاب‌سرا، چاپ نهم، ۱۳۸۸.
۱۷.م. انصاری و س. افتخار، فرهنگ نام سخن، نشر سخن، ۱۳۸۴.
۱۸.ع. بهنیا، پژوهشی در نام‌های ایرانیان معاصر، انتشارات مهدیه، چاپ دوم، ۱۳۶۳.
۱۹.م. ر. اکبری بیرقی، دایره‌المعارف نامهٔ نشر اردیبهشت، ۱۳۷۱.
۲۰.علی‌اکبر دهخدا، لغت‌نامه، لوح فشرده.
۲۱.محمد معین، فرهنگ فارسی معین، لوح فشرده.
۲۲.فرهنگ واژه‌های اوستا، احسان بهرامی، نشر بلخ، چاپ یکم، ۱۳۶۹.
۲۳.کریستین بارتلمه، تاریخچه واج‌های ایرانی، ترجمهٔ واهه دومانیان، انتشارات پژوهشگاه علوم انسانی و مطالعات فرهنگی، چاپ نخست، ۱۳۸۴.
۲۴.منوچهر آریان‌پور کاشانی، فرهنگ ریشه‌های هندواروپایی زبان فارسی، انتشارات جهاد دانشگاهی اصفهان، چاپ نخست، ۱۳۸۴.
۲۵.رودریگر اشمیت، راهنمای زبان‌های ایرانی، ترجمه آرمان بختیاری و همکاران، جلد ۱ و ۲، انتشارات ققنوس، چاپ نخست، ۱۳۸۲.
۲۶.تقی وحیدیان کامیار و غلامرضا عمرانی، دستور زبان فارسی (۱)، انتشارات سمت، چاپ یازدهم، ۱۳۸۷.
۲۷.یدالله ثمره، آواشناسی زبان فارسی، آواها و ساخت آوایی هجا، مرکز نشر دانشگاهی، ویرایش دوم، چاپ نهم، ۱۳۸۳.
۲۸.ایران کلباسی، ساخت اشتقاقی واژه در فارسی امروز، پژوهشگاه علوم انسانی و مطالعات فرهنگی، چاپ سوم، ۱۳۸۷.
۲۹.محسن ابوالقاسمی، تاریخ زبان فارسی، انتشارات سمت، چاپ نخست، ۱۳۷۳.
۳۰.ضیاءالدین هاجری، فرهنگ وندهای زبان فارسی، انتشارات آوای نور، چاپ یکم، ۱۳۷۷.
۳۱.تقی وحیدیان کامیار، عبدالحسین زرین‌کوب و حمید زرین‌کوب، ادبیات فارسی (قافیه و عروض - نقد ادبی)، شرکت چاپ و نشر کتاب‌های درسی ایران، چاپ چهارم، ۱۳۷۷.
۳۲.ا. جی. راسل و پ. نورویگ، هوش مصنوعی: مفاهیم بنیادین، ترجمهٔ س. راحتی، م. بهداد و ح. تیموری، انتشارات دانشگاه امام رضا، چاپ دهم، ۱۳۸۹.
34.http://www.britannica.com
35. http://www.sabteahval.ir
36.G. Cybenco, Approximation by Superpositions of a Sigmoidal Function, Mathematics of Control, Signals and Systems, vol. 2, 1989.
37.k. M. Hornik, M. Stinchcombe and H. White, Multilayer FeedForward Networks are Universal Approximators, Neural Networks, vol. 2, 1989.

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026

**پیوست ۱) نمونه‌ای از فرم‌های نام‌های جامعهٔ آغازین، فرستاده شده برای جامعهٔ آماری**

| ردیف | نام | معنا | زیبایی معنایی | زیبایی آوایی |
|---|---|---|---|---|
| ۱ | اردشیر | پادشاهی مقدس / سرزمین | | |
| ۲ | اردوان | پاسدار راستی / نگهبان | | |
| ۳ | ارژنگ | جادویی / زینت / نقش و نگار | | |
| ۴ | البرز | کوه بلند و بزرگ/ از همه‌سو | | |
| ۵ | الوند | دارای تندی و تیزی / تندمند | | |
| ۶ | اندیشه | فکر / خرد / بررسی همه‌جانبه | | |
| ۷ | ایزدیار | آن که خداوند یار اوست / دوست | | |
| ۸ | آرشیدا | خورشید آریایی / ایرانی درخشان | | |
| ۹ | بهنام | خوشنام / دارای نام نیک | | |
| ۱۰ | بیژن | شجاع / جنگ‌جو | | |
| ۱۱ | پردیس | فردوس / بهشت / بستان | | |
| ۱۲ | پرگل | برگ گل | | |
| ۱۳ | پولاد | فولاد / آهن آبدیده | | |
| ۱۴ | تهمتن | دارنده تن نیرومند و بی همتا / | | |
| ۱۵ | خجسته | فرخنده / مبارک / نیک | | |
| ۱۶ | دادبه | عادل / صاحب بهترین عدل | | |
| ۱۷ | دریا | دریا | | |
| ۱۸ | رستا | رهنده / کامیاب / رستگار | | |
| ۱۹ | رودابه | داده بهشت / دارای فرزند تابان | | |
| ۲۰ | روزبه | بهروز / نیکبخت | | |
| ۲۱ | سپیده | سحرگاه / نخستین پرتو خورشید | | |
| ۲۲ | سمن | یاسمن / چهره سفید و لطیف | | |
| ۲۳ | سوگند | قسم / پیمان | | |
| ۲۴ | شیوا | فصیح / بلیغ / رسا | | |
| ۲۵ | فرانک | پروانه | | |
| ۲۶ | فرزانه | عاقل / حکیم / دانا | | |
| ۲۷ | فرناز | دارای ناز زیبا / برازنده | | |
| ۲۸ | فرنگیز | دارای گیسوان نرم و فراوان | | |
| ۲۹ | فرهود | درست دین / راست آیین | | |
| ۳۰ | فروتن | متواضع / افتاده | | |
| ۳۱ | کیهان | جهان / هستی / گیتی | | |
| ۳۲ | گلفام | به رنگ گل / گلگون | | |
| ۳۳ | مروارید | جواهر گرانبها / سنگ قیمتی | | |
| ۳۴ | منوچهر | بهشت روی / دارای چهره مینوی | | |
| ۳۵ | منیژه | بانوی دوست داشتنی | | |
| ۳۶ | مهتاب | پرتو و نور ماه | | |
| ۳۷ | مهرآرا | آراینده مهر | | |
| ۳۸ | مینا | آبگینه / پرنده زیبای سخنگو | | |

| | | | | |
|---|---|---|---|---|
| ۳۹ | هستی | وجود / زندگی / آفرینش / دارایی | | |
| ۴۰ | همایون | خجسته / فرخنده / دارای تاثیر | | |
| ۴۱ | هنگامه | غوغا / معرکه / آشوب | | |
| ۴۲ | هومن | اندیشه نیک / نیک منش | | |

**واژنامه**

1. Evolutionary computing
2. Genetic Algorithm (GA)
3. Artificial Neural Network (ANN)
4. onomastics
5. Aryan
6. Indo-European
7. Frantz Bopp
8. Indo-German
9. morphology
10. vowel gradation / ablaut
11. Jakob Ludwig Carl Grimm
12. cross-over
13. Matlab Toolbox Ver. 7.8
14. feed-forward back-propagation
15. Levenberg-Marquardt
16. Multi-Layer Perceptron (MLP)

Downloaded from jor.iranaict.ir at 18:36 +0330 on Sunday November 22nd 2026