Islamic Azad University , Shiraz Branch
نشریه تحلیل مدارها، داده ها و سامانه ها
Journal of Circuits, Data and Systems Analysis
sanad.iau.ir/journal/jcdsa

Automatic recognition of digital speech using deep spiking neural network based on fuzzy weighting

Melika Hamian[1], Karim Faez[*2], Sohila Nazari[3], Maliheh Sabeti[4]

[1] Department of Computer Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran.
hamian.melika@gmail.com
[2] Department of Computer Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran.
karim.faeez@gmail.com
[3] Department of Computer Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran.
msoheilanazari21@yahoo.com
[4] Department of Computer Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran.
malihe.sabeti@gmail.com

Abstract: Despite the progress made in the design of spiking neural networks (SNN), training these systems for classification and artificial intelligence applications is one of the upcoming challenges for their design. In this paper, we have investigated supervised learning in SNNs for the problem of digit recognition and classification from speech signals. SNN training is done using fuzzy logic. In this method, the learning rule integrates Fuzzy Weighting System (FWS) with Spike Time Dependent Flexibility (STDP). SNN uses a set of training neurons with fuzzy weighting to reduce the number of weights of each neuron in the training phase, in which the data related to all classes are fed to these neurons to determine the training weights and threshold estimation with the help of the Wild Horse Algorithm (WHO). Then, these rule weights are given to the neurons of different layers to reflect the similarities in the extracted features among the classes as an objective function. A case study has been carried out on a set of audio signal data for digit classification. Our network achieved a classification accuracy of 98.17% on the TIDIGITS test database.

Keywords: digit speech recognition, spiking neural network (SNN), fuzzy weighting system (FWS)

JCDSA, Vol. 2, No. 6, Summer 2024	Online ISSN: 2981-1295	Journal Homepage: https://sanad.iau.ir/en/Journal/jcdsa
Received: 2023-12-08	Accepted: 2024-05-29	Published: 2024-09-14

CITATION

Hamian, M., et. al. *"Automatic recognition of digital speech using deep spiking neural network based on fuzzy weighting"*, Journal of Circuits, Data and Systems Analysis (JCDSA), Vol. 2, No. 6, pp. 10-18, 2024.
DOI: 00.00000/0000

COPYRIGHTS

©2024 by the authors. Published by the Islamic Azad University Shiraz Branch. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0)
https://creativecommons.org/licenses/by/4.0

[*] Corresponding author

Extended Abstract

1- Introduction

In this work, we have used a series of meta-heuristic algorithms for deep neural network training based on Fuzzy Weighted System (FWS) learning rule integrated with Spike Time Dependent Flexibility (STDP). Our case study is for classification and recognition of digits from audio signals. The important innovation of this training is the use of a random weighting system based on fuzzy logic (FWS), which is sufficient for each neuron. Instead of defining a large volume of input weights for each neuron, only two variables are calculated with the help of the WHO algorithm. A training audio dataset is then trained for the entire network applying audio samples to these compressed spike maps with trained weights, which are then transformed into the original classification results using a spiking autoencoder. To the best of our knowledge, this is the first work that performs audio synthesis from extracted features in a spike-based environment with the help of optimization algorithms and a fuzzy logic system.

2- Methodology

To train FWS-DSNN-based deep voice models, which is the main contribution of this work, several popular speech features are extracted from the training recordings provided for digit classification. After this, with the help of the selected features, we perform the steps of FWS-DSNN training. For each feature in the input layer of the network with the Izhikevich model, a spike train of different features is generated. Now it's time to define the weight and threshold based on the training data. These data enable the training of the FWS-DSNN deep audio model with a back-to-back learning approach. During training, FWS-DSNN learns to optimize selected input speech features by fuzzy weighting and threshold selection optimized with the WHO Wild Horse optimization algorithm with input spike trains from multiple intermediate and output layers of spiking neurons.

We have used a mathematical function A.cos(t.s) for random weighting. This work means that for a multi-layer network with multiple neurons in each layer, we no longer need to calculate and estimate many weights to define each neuron and calculate only two parameters Z1, Z2. This helps to reduce the parameters that can be calculated to define the neurons in each layer. This parameter reduction helps to simplify the learning and training of FWS-DSNN. In this regard, we have used the wild horse optimization algorithm to optimize the calculation of Z1, Z2 parameters along with the threshold value (Vth). Through a fuzzy weighted layering of different neurons in different layers, we get a random distribution of weights and threshold value, which leads to the training of the FWS-DSNN network. During the inference stage, the audio digit provided by the trained FWS-DSNN model is combined with the information stored in the language model and pronunciation vocabulary.

3- Results and discussion

In this article, the wild horse algorithm is used to train the proposed network. Therefore, the proposed FWS-SNN method with the WHO algorithm approach has been compared with two machine learning methods, which include feedforward ANN method and adaptive neural fuzzy network ANFIS for the classification of audio signals for digits 0-9. These methods are reviewed for the test data set of 30 members based on the training data of 120 members trained under different techniques. For this case study, the defined network of a three-layer network with the number of neurons [15 8] has been used for all machine learning networks. The proposed technique with the help of meta-heuristic algorithms has been able to produce good results of high accuracy and alignment of results. The WHO-FWS-SNN technique has been able to create the highest accuracy compared to other machine learning methods and achieve 98% accuracy for audio data.

4- Conclusion

In this paper, a new sequential learning neural classifier for spiking neural network, called FWS-SNN, is presented for digit classification problems from audio signals. The learning algorithm uses low computational cost weight update rules that require only two parameters for each neuron. This method reduces the complexity of training the studied neural network. We also describe a number of strategies for optimizing SNNs for implementation on memory- and energy-constrained hardware, including approximations in computing neural dynamics and reduced precision in storing synaptic weights. The fuzzy weighting technique has been able to reduce the complexity of the proposed SNN network by reducing 71.9% of the network parameters for neuron weighting.
As future work, we will investigate the training of recurrent networks of spiking neurons for speech recognition applications for digit classification to improve recognition performance. For this purpose, a second type of fuzzy modeling can be used to define neurons to increase accuracy.

Islamic Azad University , Shiraz Branch
نشریه تحلیل مدارها، داده ها و سامانه ها
Journal of Circuits, Data and Systems Analysis
sanad.iau.ir/journal/jcdsa

تشخیص خودکار گفتار رقمی با استفاده از شبکه عصبی اسپایکینگ عمیق بر اساس وزن دهی فازی

ملیکا حامیان[۱]، کریم فایز[۲*]، سهیلا نظری[۳]، ملیحه ثابتی[۴]

۱- گروه مهندسی کامپیوتر، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران (Hamian.melika@gmail.com)
۲- گروه مهندسی کامپیوتر، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران (karim.faeez@gmail.com)
۳- گروه مهندسی کامپیوتر، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران (msoheilanazari21@yahoo.co)
۴- گروه مهندسی کامپیوتر، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران (malihe.sabeti@gmail.com)

چکیده: علی‌رغم پیشرفت‌های انجام شده در طراحی شبکه‌های عصبی اسپایکینگ، آموزش این سیستم‌ها برای طبقه‌بندی و کاربردهای هوش مصنوعی از چالش‌های پیش‌رو برای طراحی آن‌هاست. در این مقاله ما یادگیری نظارت‌شده را در شبکه‌های عصبی اسپایکی برای مساله تشخیص و طبقه‌بندی رقم از روی سیگنال های گفتار، بررسی کرده‌ایم. در این روش، قانون یادگیری سیستم وزن‌دهی فازی با انعطاف‌پذیری وابسته به زمان اسپایک ادغام می‌شوند. قانون انعطاف‌پذیری وابسته به زمان اسپایک ترکیب شده با سیستم وزن‌دهی فازی، توزیع وزن تصادفی را ایجاد می‌کند که در آن محدوده پنجره انعطاف‌پذیری وابسته به زمان اسپایک کنترل می‌شود. شبکه عصبی اسپایکینگ از یک مجموعه نورون آموزشی با وزن‌دهی فازی برای کاهش تعداد وزن‌های هر نورون، در مرحله آموزش استفاده می‌کند که در آن داده‌های مرتبط با تمام کلاس‌ها به این نورون‌ها جهت تعیین وزن‌های آموزش و تخمین آستانه با کمک الگوریتم اسب وحشی، اعمال می‌شود. سپس این قانون وزن‌ها، به نورون‌های لایه‌های مختلف داده می‌شوند تا شباهت‌ها را در ویژگی‌های استخراج شده در بین کلاس‌ها به عنوان تابع هدف، منعکس نماید. نتایج روش پیشنهادی، دقت طبقه‌بندی ۹۸٪/۱۷ در پایگاه داده آزمایشی TIDIGITS را نشان می‌دهد.

واژه های کلیدی: سیستم تشخیص ارقام، شبکه عصبی اسپایکینگ، سیستم وزن‌دهی فازی

نوع مقاله: پژوهشی
DOI: 00.00000/0000
تاریخ ارسال مقاله: ۱۴۰۲/۰۹/۱۹ تاریخ پذیرش مقاله: ۱۴۰۳/۰۳/۰۶ تاریخ چاپ مقاله: ۱۴۰۳/۰۶/۲۲

۱- مقدمه

در طی چند سال اخیر محققان حوزه پردازش گفتار تلاش‌های زیادی برای بهبود عملکرد سیستم‌های خودکار بازشناسی گفتار[۱] در شرایط تمیز انجام داده‌اند. مقاوم‌سازی سیستم بازشناسی نسبت به تنوعات مختلف گفتاری (مانند تنوعات گوینده، لهجه، نویز محیط، کلنال انتقال و ...) نیز از دیگر حوزه‌های فعال در بحث بازشناسی گفتار است [۱]. بیشتر تحقیقات انجام شده در زمینه مقاوم‌سازی بازشناسی گفتار نسبت به تنوعات، روی سه روش عمده بهسازی گفتار، استخراج ویژگی‌های مقاوم و جبران سازی پارامترهای مدل صوتی متمرکز شده است [۲]. از طرفی دیگر، تحقیقات اخیر نشان می‌دهد که نتایج به دست آمده از بهترین سیستم‌های خودکار بازشناسی گفتار، پایین‌تر از نتایج بازشناسی سیستم شنوایی انسان است. از این رو، می‌توان امید داشت با الهام گرفتن از عملکرد فیزیولوژیک شنوایی انسان، بازشناسی این گونه سیستم‌ها را افزایش داد [۳]. تشخیص گفتار به یک راهکار مهم برای بهبود رابط انسان و ماشین تبدیل شده است. با در نظر گرفتن محدودیت‌های سیستم‌های تشخیص خودکار گفتار فعلی (مانند راه‌حل‌های غیرواقعی مبتنی بر ابر)، علاقه به شبکه‌های عصبی و سیستم‌های الهام‌گرفته از زیستی، انگیزه اجرای روش‌های جدید را فراهم کرده است. شبکه‌های عصبی مصنوعی[۲] به روش اصلی مدل‌سازی صوتی برای تشخیص خودکار گفتار واژگان بزرگ تبدیل شده‌اند. یک شبکه عصبی مصنوعی معمولی دارای یک معماری چند لایه است که به مقادیر زیادی محاسبات نیاز دارد و از پیچیدگی سخت افزاری رنج می‌برد [۴]. پردازش اطلاعات در مغز انسان از طریق فعال‌سازی نورون‌های حسی و متعاقباً ارسال ورودی‌ها به نورون‌های قشر مغز انجام می‌شود که منجر به الگوهای پیچشی پیچیده جمعیت‌های عصبی برای تصمیم‌گیری یا ذخیره اطلاعات می‌شود [۵، ۶].

[1] Automatic Speech Recognition (ASR)

[2] Artificial Neural Network (ANN)

در سال‌های اخیر، پیشرفت‌هایی در الگوریتم‌های الهام‌گرفته از علوم اعصاب با توسعه شبکه‌های عصبی اسپایکینگ[1] انجام شده است [۷، ۸]. این شبکه‌ها که از مدل‌های دینامیکی مختلف نورون‌های بیولوژیکی استفاده می‌کنند، مدل‌های محاسباتی هستند که اطلاعات را در حوزه زمان رمزگذاری و پردازش می کنند [۹]. شبکه‌های عصبی عمیق، شبکه‌های عصبی مصنوعی متشکل از بیش از دو لایه عصبی هستند که از آنالوگ ساختاری و عملکردی شبکه‌های قشری بسیار ساده‌شده، شکل گرفته اند [۱۰]. علاوه بر این، مدل‌های یادگیری عمیق می‌توانند در مورد اینکه چگونه مغز ممکن است به کارهای پیچیده در محیط‌های تعریف نشده دست یابد، توضیحات و مفروضاتی ارائه دهند [۱۱]. مدل‌های یادگیری عمیق معمولاً در بسیاری از انواع داده‌ها عملکرد خوبی دارند، اما همواره یک مجموعه داده بزرگ برای آموزش آنها برای تولید نتایج معنی‌دار ضروری است [۱۰]. با توجه به تعاریف انجام شده برای شبکه عصبی اسپایکینگ و قابلیت اجرای سخت افزاری ساده این سیستم در پردازنده‌های کم مصرف، به کاربردهای مهم در این شبکه‌ها برای طبقه‌بندی و پردازش‌های هوش‌مصنوعی دست یافته‌ایم.

در این مقاله، از یک سری الگوریتم‌های فراابتکاری برای آموزش شبکه عصبی عمیق بر اساس قانون یادگیری سیستم وزن‌دهی فازی[2] ادغام شده با انعطاف پذیری وابسته به زمان اسپایک[3] استفاده شده است. مطالعه موردی ما برای کلاسه بندی و تشخیص ارقام از روی سیگنال‌های صوتی می‌باشد. در این راستا از یک الگوریتم بهینه‌سازی گرگ خاکستری[4] برای انتخاب ویژگی‌های صوتی مقاوم در برابر تغییرات محیطی و نویزهای همراه، استفاده شده است. در یک محیط مبتنی بر اسپایک، ویژگی‌های انتخابی به‌عنوان ورودی را می‌توان به نگاشت‌های اسپایک مکانی-زمانی فشرده تبدیل کرد؛ که سپس می‌توان در مدل‌های شبکه و روش‌های داده منتقل شود. رمزگذارهای خودکار اسپایکینگ براساس مدل نورونی ایزیکویچ[5] برای طبقه ورودی استفاده شده و در مجموعه داده‌های TIDIGITS آموزش و آزمایش شد. همچنین یک چارچوب سنتز صدا از روی ویژگی‌های رمزشده، متشکل از شبکه‌های عصبی اسپایکینگ چند لایه و کاملاً متصل ارائه شد. سپس برای تولید نگاشت‌های اسپایک فشرده فضایی-زمانی، ویژگی‌های انتخابی از نورون‌های مدل نشت‌یافته ادغام و آتش[6] استفاده می‌شود. برای آموزش این نورون‌ها و به منظور تعیین وزن‌های و ولتاژ آستانه نورون‌های تعریف شده در شبکه از الگوریتم اسب وحشی[7] استفاده شده است. اما نوآوری مهم این آموزش استفاده از یک سیستم وزن‌دهی تصادفی مبتنی بر منطق فازی است که برای هر نورون کافی است به جای تعریف حجم زیادی از وزن‌های ورودی هر نورون، تنها دو متغیر با کمک الگوریتم اسب وحشی محاسبه شود. سپس یک مجموعه داده‌های صوتی آموزشی برای کل شبکه می‌آموزد که نمونه‌های صوتی را به این نگاشت‌های فشرده شده اسپایک با وزن‌های آموزش داده شده اعمال کند. سپس با استفاده از رمزگذار خودکار اسپیکینگ به نتایج طبقه‌بندی اصلی تبدیل می‌شوند. تا جایی که ما می‌دانیم، این اولین کاری است که سنتز صدا از ویژگی‌های استخراج شده را در یک محیط مبتنی بر اسپایک با کمک الگوریتم‌های بهینه‌سازی و سیستم منطق فازی انجام می‌دهد.

۲- کارهای مرتبط

آموزش شبکه‌های عصبی اسپایکینگ عمیق[8] کار ساده‌ای نیست. در [۱۲]، یک چارچوب جدید تبدیل شبکه عصبی مصنوعی به شبکه عصبی اسپایکینگ و چارچوب یادگیری لایه‌ای را برای تشخیص الگوی سریع و کارآمد پیشنهاد می‌کند که به عنوان یادگیری پیشرونده شبکه‌های عصبی اسپایکینگ عمیق شناخته می‌شود. با مطالعه هم‌ارزی بین شبکه عصبی مصنوعی و شبکه عصبی اسپایکینگ در فضای نمایش گسسته، یک روش تبدیل شبکه بدوی معرفی می‌شود که از حداکثر استفاده از تعداد اسپیک برای تقریب مقدار فعال‌سازی نورون‌های آنالوگ استفاده می‌کند. برای جبران خطاهای تقریبی ناشی از تبدیل شبکه اولیه، یک روش یادگیری لایه‌ای را با یک زمان‌بندی آموزش تطبیقی برای تنظیم دقیق وزن‌های شبکه معرفی می‌کند. مقاله [۱۳] یک تحلیل مبتنی بر شبیه‌سازی برای طبقه‌بندی صدای پای انسان در محیط طبیعی با استفاده از ویژگی‌های حوزه زمان ساده انجام داد. در این تحلیل از شبکه عصبی اسپایکینگ، شامل یک طبقه‌بندی محاسباتی کم وزن که مشتق شده از یک شبکه عصبی مصنوعی است، برای طبقه‌بندی صداهای آکوستیک استفاده شد. رمزگذاری کارآمد از سیگنال‌های واقعی به اسپایک‌ها بسیار مهم است و به طور قابل توجهی بر عملکرد کلی سیستم تأثیر می‌گذارد. برای رمزگذاری موثر سیگنال‌ها به اسپایک‌ها، هم حفظ اطلاعات مربوط به کار مورد نظر و هم چگالی اسپایک‌های کدگذاری شده باید در نظر گرفته شود. در [۱۴]، چهار روش رمزگذاری اسپایک را در زمینه یک سیستم طبقه‌بندی رقمی مستقل از بلندگو مطالعه می‌کند. این روش‌ها دقت طبقه‌بندی را بهبود می‌دهند. استفاده از ترکیب شبکه‌های عصبی پراکنده و حسگرهای شنوایی نورومورفیک جایگزینی برای انجام وظیفه پردازش گفتار شبیه انسان ارائه می‌دهند. در [۱۵]، یک مدل شبکه عصبی کانولوشن اسپایکینگ پیاده‌سازی شد که در آن وزن اتصالات با آموزش یک شبکه عصبی کانولوشن با توابع فعال‌سازی خاص، با استفاده از تصاویر استاتیک مبتنی بر سرعت شلیک با اطلاعات اسپایکینگ به‌دست‌آمده از حلزون عصبی محاسبه شده است. این سیستم با یک مجموعه داده بزرگ که حاوی دستورات گفتاری «چپ» و «راست» است، آموزش داده و آزمایش شد و دقت ۸۹/۹۰٪ را به دست آورد.

[1] Spiking Neural Networks (SNN)
[2] Fuzzy Weighting System (FWS)
[3] Spike Time Dependent Flexibility (STDP)
[4] Grey Wolf Optimization (GWO)
[5] Izhikevich neuron model
[6] leaky integrate-and-fire (LIF)
[7] Wild Horse Algorithm (WHO)
[8] Deep Spiking Neural Networks (DSNN)

در [۵]، شبکه‌های عصبی اسپایکینگ را با پویایی بازگشت ذاتی بهبود یافته پیشنهاد کرده است که قادر به یادگیری موثر دنباله‌های طولانی هستند. مزیت معماری‌های پیشنهادی کاهش تعداد پارامترهای قابل آموزش در مقایسه با حافظه طولانی کوتاه مدت[1] است. طرح آموزشی ارائه شده برای آموزش معماری‌های پیشنهادی به شبکه‌های عصبی اسپایکینگ اجازه می‌دهد تا خروجی‌های چند بیتی (برخلاف اسپایک‌های باینری ساده) تولید کنند و به مشکل عدم تطابق گرادیان که به دلیل استفاده از تابع جایگزین برای غلبه بر عدم تمایز نورون‌های اسپکینگ رخ می‌دهد، کمک کند. در [۱۶]، یک شبکه عصبی اسپایکینگ سلسله مراتبی[2] بهینه‌سازی شده برای به حداکثر رساندن دقت تشخیص کلمه در محیط پرسروصدا و چندین سخنران سلسله مراتب سازمانی مسیر شنوایی صعودی را پیش‌بینی می‌کند. مقایسه با داده‌های مربوط به عصب شنوایی، مغز میانی، تالاموس و قشر مغز نشان می‌دهد که شبکه عصبی اسپایکینگ سلسله مراتبی بهینه چندین تغییر مسیر شنوایی صعودی از جمله تلفات متوالی قدرت تفکیک زمانی و تولنایی هماهنگ‌سازی، افزایش پراکندگی و انتخاب‌پذیری را پیش‌بینی می‌کند. در [۱۷]، یک طرح رمزگذاری و رمزگشایی عصبی اسپایکینگ را پیشنهاد می‌کند که برای پردازش صدا بهینه شده است.

در [۱۸]، روشی را برای سنتز تصاویر از روش‌های مختلف در یک محیط مبتنی بر سنبله پیشنهاد می‌کند. در این کار از رمزگذارهای خودکار اسپایکینگ برای تبدیل ورودی‌های تصویر و صدا به نمایش‌های فشرده مکانی-زمانی استفاده شد. در [۱۹] یک چارچوب قابل قبول بیولوژیکی، یعنی شبکه عصبی اسپایکینگ با خودسازماندهی بدون نظارت را پیشنهاد می‌کند. این چارچوب از نقشه خودسازماندهی بدون نظارت[3] برای نشان دادن محتویات فرکانس تعبیه شده در سیگنال‌های صوتی استفاده می‌کند و به دنبال آن از یک شبکه عصبی اسپایکینگ مبتنی بر رویداد برای طبقه‌بندی الگوی اسپایکینگ فضایی-زمانی استفاده می‌کند. نتایج تجربی در مورد پایگاه داده صدای محیطی RWCP و مجموعه داده‌های رقم گفتاری TIDIGITS انجام شد، که دقت طبقه‌بندی رقابتی را نسبت به سایر مدل‌های مبتنی بر یادگیری عمیق و شبکه عصبی اسپایکینگ معمولی نشان می‌دهد. در [۲۰]، از شبکه عصبی اسپایکینگ برای مدل‌سازی صوتی استفاده می‌کند و عملکرد آن‌ها را در چندین سناریو تشخیص واژگان بزرگ ارزیابی می‌کند. نتایج تجربی دقت خودکار بازشناسی گفتار رقابتی را به همتایان شبکه عصبی مصنوعی خود نشان می‌دهند؛ در حالی که برای طبقه‌بندی هر فریم صوتی تنها به ۱۰ مرحله زمانی الگوریتمی و به ۰/۶۸ برابر کل عملیات سیناپسی نیاز دارند.

با مطالعه تکنیک‌های مختلف اشاره شده در این بخش می‌توان به این نتیجه رسید که رویکردهای مختلفی برای آموزش و تعریف شبکه‌های عصبی اسپایکینگ ارائه شده است. در این مقاله با توجه به این مطالعات انجام شده سعی شد یک راه‌کار نوین و ساده برای نحوه آموزش شبکه عصبی اسپایکینگ ارائه گردد تا ضمن کاهش پیچیدگی‌های آموزش به یک راه‌کار برای افزایش دقت آموزش دست یابد.

۳- مفاهیم مقدماتی

۳-۱- سیستم منطق فازی

مساله مهم در دسته‌بندی مجموعه اعداد این است که به طور قطع نمی‌توان گفت یک داده به صورت صددرصد به یک کلاس تطبیق می‌یابد. بنابراین مفهوم مجموعه فازی تعریف شده تا بتوان اعداد را بر اساس یک مدل تابع عضویت در یک مجموعه خاص به صورت نسبی اختصاص داد. مهمترین مرحله در پردازش فازی به پردازش اطلاعات با تکنیک فازی انتخاب شده و مشکلی که باید حل شود بستگی دارد. پردازش فازی دارای سه مرحله اصلی فازی‌سازی اطلاعات، استنتاج فازی بر اساس قوانین فازی تعریف شده، فاززدایی اطلاعات است. مراحل فازی‌سازی و فازی‌زدایی شامل کدگذاری داده‌ها (فازی‌سازی) و رمزگشایی نتایج (فازی‌زدایی) است. این مراحل پردازش فازی را با روش فازی امکان پذیر می‌کند. بنابراین، فازی‌سازی و فازی‌زدایی مهم‌ترین مراحلی هستند که تولنایی مدیریت مساله با تکنیک‌های قوانین فازی را دارا است. دلیل اینکه منطق فازی بهتر از سایرین عمل می‌کند، این است که همه چیز از عدم دقت رنج می‌برد؛ در حالیکه منطق فازی درک خود را با در نظر گرفتن ساختار می‌سازد [۲۱]. از سوی دیگر، دلایل بسیاری مانند تصادفی بودن و ابهام منجر به عدم قطعیت در نتیجه پردازش اطلاعات و داده‌ها می‌شود [۲۲، ۲۳]. بر این اساس در این مقاله از یک الگوی فازی تصادفی برای تعریف وزن‌های مورد نیاز برای نورون‌های شبکه عصبی اسپایکینگ بهره خواهیم برد.

۳-۲- شبکه عصبی اسپایکینگ

شبکه‌های عصبی اسپایکینگ از تعداد زیادی عنصر پردازشی فوق‌العاده به‌هم‌پیوسته به نام نورون (با مدلسازی عصب‌های مغز) تشکیل‌شده که برای حل یک مسئله با یکدیگر به‌صورت هماهنگ عمل می‌کنند. این نورون‌ها با کمک اسپایک به هم ارتباط برقرار می‌کنند. شبکه عصبی شبکه‌ای از لایه‌هاست؛ معمولاً لایه‌ای که اطلاعات ورودی به آن‌ها داده می‌شود تحت عنوان لایه‌ی ورودی و لایه‌ای که داده‌های خروجی از آن دریافت می‌شود تحت عنوان لایه‌ی خروجی نامیده می‌شود و به لایه‌های دیگر بین این دولایه (در صورت وجود) لایه‌های پنهان گفته می‌شود. برای تعداد لایه های پنهان بالا، شبکه عصبی اسپایکینگ عمیق تشکیل می‌شود. شبکه‌های عصبی اسپایکینگ، سیستم‌های محاسباتی هستند که توسط شبکه‌های عصبی زیستی الهام گرفته شده‌اند که مغز حیوانات را تشکیل می‌دهند. منظور از یادگیری در شبکه‌های عصبی، تنظیم وزن‌ها و آستانه‌گذاری نورون‌ها در شبکه

[1] Long short-term memory (LSTM)
[2] Hierarchical Spiking Neural Networks (HSNN)
[3] Self-Organizing Map (SOM)

می‌باشد. بر این اساس الگوریتم‌های متفاوتی بیان شده، که معمول‌ترین آنها یادگیری دلتا، یادگیری رقابتی و انعطاف پذیری وابسته به زمان اسپایک می‌باشد.

۳-۳- الگوریتم بهینه سازی گرگ خاکستری

الگوریتم گرگ خاکستری توسط سید علی میرجلیلی در [۲۴] سال ۲۰۱۴ ارائه شده است. این الگوریتم فراکتشافی یا الگوریتم بهینه‌سازی بر اساس رفتار و شیوه شکار گرگ‌های خاکستری است. این الگوریتم مبتنی بر جمعیت بوده و فرآیند ساده‌ای دارد و به سادگی قابلیت تعمیم به مسائل با ابعاد بزرگ را دارد.

۴-۳- الگوریتم بهینه‌سازی اسب وحشی

الگوریتم‌های بهینه‌سازی معمولاً از رفتار طبیعی یک عامل الهام می‌گیرند که می‌تواند انسان، حیوان، گیاه یا یک عامل فیزیکی یا شیمیایی باشد. بسیاری از الگوریتم‌های ارائه شده در دهه گذشته از رفتار حیوانات الهام گرفته شده‌اند. در این مقاله از یک الگوریتم بهینه‌سازی جدید به نام بهینه‌ساز اسب وحشی استفاده می‌کنیم که از رفتار اجتماعی اسب‌های وحشی الهام گرفته شده است. اسب‌ها معمولاً در گروه‌هایی متشکل از یک اسب نر و چند مادیان و کره اسب زندگی می‌کنند. اسب‌ها رفتارهای زیادی از خود نشان می‌دهند؛ مانند چرا، تعقیب، تسلط، رهبری و جفت‌گیری. رفتار جذابی که اسب‌ها را از سایر حیوانات متمایز می‌کند، ادب آن‌هاست. رفتار پرورش اسب به گونه‌ای است که کره‌اسب‌ها قبل از رسیدن به سن بلوغ گروه را ترک کرده و به گروه‌های دیگر می‌پیوندند. این خروج برای جلوگیری از جفت شدن پدر با دختر یا خواهر و برادر است. الهام‌بخش اصلی الگوریتم پیشنهادی رفتار مؤدبانه اسب است [۲۵].

۴- روش یادگیری پیشنهادی

چالش مهم در مساله یادگیری شبکه‌های عصبی اسپایکینگ عمیق، پیچیدگی‌های بسیار بالای تعیین مقادیر آستانه و وزن‌های هر نورون برای تعداد لایه‌های بالا و افزایش تعداد نورون‌ها می‌باشد. در این طرح پیشنهادی تلاش شده با کمک یک سیستم وزن دهی فازی این مقدار را کاهش دهیم. شکل (۱) نمای کلی رویکرد آموزش شبکه عصبی اسپایکینگ عمیق را نمایش داده است. در ادامه به تفسیر کامل این چارچوب پیشنهادی خواهیم پرداخت.

۱-۴- مدل نورون اسپکینگ

همانطور که در شکل (۱) نشان داده شده، ابتدا ویژگی های مبتنی بر فریم استخراج شده و به مدل های آکوستیک مبتنی بر شبکه‌های عصبی اسپایکینگ وارد می‌شوند. با توجه به مدت زمان کوتاه فریم‌های قطعه‌بندی شده و تغییرات آهسته سیگنال‌های گفتاری، این ویژگی‌ها معمولاً در طول دوره زمانی کوتاه فریم‌های قطعه‌بندی شده، ثابت فرض می‌شوند. مطابق با مقادیر ثابت برای ویژگی‌های استخراج شده ما برای تولید قطار اسپایک از مدل ایزیکویچ [۲۶] برای نورون‌های لایه ورودی و از مدل نورون ادغام و آتش با طرح تولید قطار پالس در نورون‌های لایه میانی و خروجی استفاده می‌شود که می‌تواند به طور موثر این ویژگی‌های مبتنی بر قاب ثابت را با حداقل هزینه‌های محاسباتی پردازش کند. اگرچه نورون‌های ادغام و آتش دینامیک زمانی غنی نورون‌های بیولوژیکی را تقلید نمی‌کنند، اما برای کار با بازنمایی عصبی که در این کار استفاده می‌شود، ایده‌آل هستند؛ با این که زمان‌بندی اسپایک نقش ناچیزی دارد. اما با تعریف حداکثر فرکانس مداری نورون‌ها، می‌توان سرعت تحلیل شبکه عصبی اسپایکینگ را بالا برد [۲۷، ۲۸].

مشارکت‌های اصلی این مقاله عبارتند از:

(۱) یک مدل نورونی که از یک مکانیسم رمزگذاری مبتنی بر مدل ایزیکویچ برای تبدیل یک ورودی با ارزش واقعی به الگوهای سنبله استفاده می‌کند.

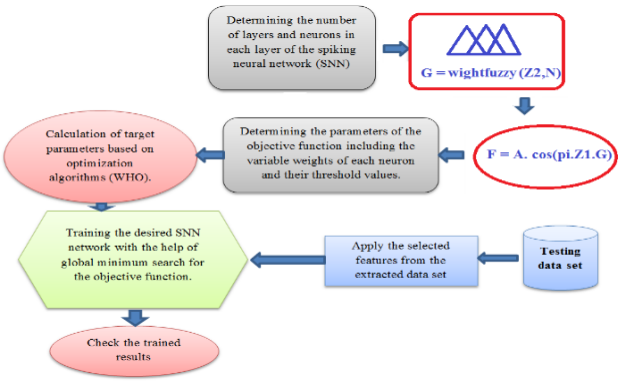
(۲) یک الگوریتم انتخاب ویژگی‌های مقاوم برای یک شبکه عصبی اسپایک عمیق که نیاز به یک پیش واحد ارسال نمونه‌های آموزشی دارد، معرفی می‌شود.

(۳) ارائه یک الگوریتم یادگیری با قابلیت تعیین خودکار معماری شبکه بر اساس نمونه‌های آموزشی است.

(۴) قوانین به‌روزرسانی وزن که به دستکاری‌های تصادفی محاسباتی با سیستم منطق فازی متکی هستند.

۲-۴- طرح آموزش شبکه عصبی

شبکه‌های عصبی اسپایکینگ اطلاعات دریافت شده از طریق قطارهای اسپایک از لایه ورودی را پردازش می‌کنند. بنابراین، مکانیسم‌های خاصی برای رمزگذاری بردارهای ویژگی سیگنال‌های صوتی با ارزش پیوسته در قطارهای اسپایک و پردازش آن‌ها در لایه‌میانی و رمزگشایی نتایج طبقه‌بندی از فعالیت نورون‌های خروجی مورد نیاز است. برای این منظور، یک طرح رمزگذاری عصبی اسپایکینگ در این مقاله پیشنهاد شده که بر اساس یک وزن‌دهی تعریف شده فازی، و جستجو با الگوریتم فرا ابتکاری انجام می‌شود. در این بخش مراحل یادگیری شبکه عصبی اسپایکینگ عمیق مبتنی بر سیستم وزن‌دهی فازی (FWS-DSNN) تشریح می‌شود.



شکل (۱): مدل آموزش FWS-DSNN

برای آموزش مدل‌های صوتی عمیق مبتنی بر FWS-DSNN، که نوآوری اصلی این کار است، چندین ویژگی گفتاری محبوب از ضبط‌های آموزشی استخراج شده که برای کلاسه‌بندی ارقام ارائه شده است. در این راستا، ما از یک مجموعه داده‌های صوتی با شرایط مختلف و نویزی با گوینده‌های مختلف مرد و زن با فرکانس ۸کیلوهرتز استفاده کرده‌ایم. ویژگی‌های استخراج شده از این سیگنال‌ها شامل ضرایب کپسترال فرکانس مل[1]، آشکارساز عبور از صفر[2] و توان سیگنال‌های صوتی به‌عنوان گزینه‌های مناسب برای طبقه بندی ارقام استفاده شده است [۳۳]. قبل از وارد شدن به FWS-DSNN، این ویژگی‌های گفتاری ورودی با اتصال فریم‌های متعدد به منظور بهره‌برداری از اطلاعات بافت زمانی بیشتر، زمینه‌سازی می‌شوند. قبل از آموزش مدل آکوستیک مبتنی بر شبکه عصبی اسپایکینگ، هم‌ترازی ویژگی‌های گفتار با برچسب‌های هدف با استفاده از سیستم انتخاب ویژگی با الگوریتم فراابتکاری بهینه‌سازی گرگ خاکستری بر اساس قوانین تعریف شده پیشنهادی انجام می‌شود. یک پراکندگی برای وزن‌های مختلف معرفی می‌شود که در این سیستم با مقدار z_1 این پراکندگی را تغییر می‌دهیم. میزان پراکندگی هر ویژگی در طبقه‌های مختلف و تراکم هر ویژگی در هر طبقه، به دست می‌آید. با تعریف این تابع، ویژگی‌هایی که نسبت به طبقه‌بندی مربوط به خود، بیشترین تراکم آن ویژگی در هر طبقه داشته باشد و بیشترین پراکندگی ویژگی نسبت به طبقات مختلف داشته باشد، به‌عنوان بهترین گزینه برای ویژگی طبقه‌بندی انتخاب می‌شود. نتایج انتخاب ویژگی با کمک الگوریتم بهینه‌سازی گرگ خاکستری انجام می شود.

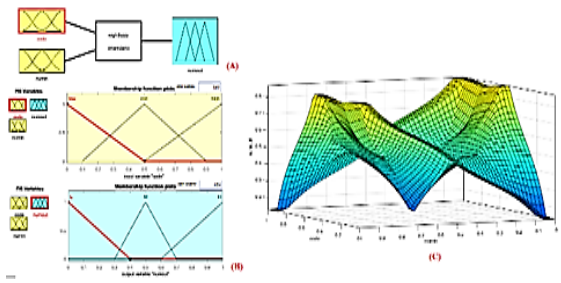
در ادامه، با کمک ویژگی‌های انتخاب شده مراحل آموزش الگوریتم FWS-DSNN انجام می‌شود. برای هر ویژگی در لایه ورودی شبکه با مدل ایزیکویچ، قطار اسپایک ویژگی‌های مختلف تولید می‌شود. حال نوبت به تعریف وزن و آستانه از روی داده‌های آموزشی است. این داده‌ها، آموزش مدل صوتی عمیق FWS-DSNN را با رویکرد یادگیری پشت سر هم امکان پذیر می‌کند. در طول آموزش، شبکه می‌آموزد که ویژگی‌های گفتار ورودی انتخاب شده را با وزن‌دهی فازی و انتخاب آستانه بهینه‌شده با الگوریتم بهینه‌سازی اسب وحشی با قطارهای اسپایک ورودی از میان لایه‌های متعدد میانی و خروجی نورون‌های اسپکینگ، بهینه کند. این طرح بهینه‌سازی فازی ابتدا بردار ویژگی ورودی مبتنی بر فریم X را تغییر می‌دهد، که در آن بردار ویژگی های $[x1, x2, \dots . xn]$ به شرح ذیل است.

$$[\mathrm{W}_1 . \mathrm{W}_2 . \dots . \mathrm{W}_N] = \text{wighting}\ (\mathrm{Z}_1 . \mathrm{Z}_2 . \dots . \mathrm{Z}_N)$$
$$F = \sum_{i=1}^{N} w_i \times x_i \qquad (۱)$$

که N تعداد ورودی‌های هر نورون است و Zi معرف پارامترهای تولید وزن بر اساس مدل فازی می‌باشد. همچنین Wi معرف وزن‌های اختصاصی هر نورون می-باشد. نکته نوآورانه این کار، وزن‌دهی چندگانه بر مبنای تولید وزن‌های تصادفی با سیستم منطق‌فازی می‌باشد. برای ایجاد یک مجموعه وزن‌دهی تصادفی از یک تابع ریاضی A.cos(t.s) بهره برده‌ایم. این کار باعث می‌شود برای یک شبکه چند لایه با نورون‌های متعدد در هر لایه ما دیگر به محاسبه و تخمین وزن‌های زیاد برای تعریف هر نورون نیاز نداشته باشیم و تنها دو پارامتر $Z1, Z2$ را محاسبه نماییم. این عمل به کاهش پارامترهای قابل محاسبه برای تعریف نورون‌ها در هر لایه کمک می‌کند. این کاهش پارامتر به ساده‌سازی یادگیری و آموزش شبکه کمک می‌کند. در این راستا برای بهینه‌سازی محاسبه پارامترهای Z_1, Z_2 به همراه مقدار آستانه از الگوریتم بهینه‌سازی اسب وحشی استفاده کرده‌ایم. از طریق یک لایه‌گذاری وزنی فازی از نورون‌های مختلف در لایه‌های مختلف به یک توزیع تصادفی از وزن‌ها و مقدار آستانه دست می‌یابیم که منجر به آموزش شبکه می‌شود. در طول مرحله استنتاج، رقم صوتی ارائه شده توسط مدل شبکه آموزش دیده با اطلاعات ذخیره شده در مدل زبان و واژگان تلفظ ترکیب می‌شود. استفاده از سیستم وزن‌دهی فازی تصادفی به عنوان یک تعریف‌کننده وزن‌های تصادفی برای ایجاد سیستم تولید وزن، به جای استفاده از تخمین همه وزن‌ها با کمک الگوریتم فراابتکاری یک روش برجسته برای کاهش پیچیدگی آموزش شبکه FWS-DSNN است. بر اساس این روش، شبکه‌های عصبی پشت سر هم آنگونه که در شکل (۱) نشان داده شده، ساخته شد. در طول انتشار فعال‌سازی رو به جلو، از لایه‌های شبکه عصبی اسپایکینگ برای تعیین نمایش دقیق سنبله استفاده می‌شود که سپس تعداد کل سنبله‌ها و قطارهای سنبله را به ترتیب، به لایه‌های FWS-DSNN بعدی منتشر می‌کند. این ساختار لایه‌درهم، اطلاعاتی را که به لایه‌های شبکه متصل شده منتشر می‌شود، هماهنگ می‌کند.

جدول (۱): نمایش قوانین فازی

code	numin	numout
Low	Low	**L**
Low	Mid	**M**
Low	High	**H**
Mid	Low	**H**
Mid	Mid	**M**
Mid	High	**L**
High	Low	**H**
High	Mid	**L**
High	High	**H**



شکل (۲): سیستم فازی تعریف شده برای ایجاد وزن‌های تصادفی (الف): نمای کلی سیستم فازی (ب): توابع عضویت ورودی و خروجی (ج): مشخصه ورودی و خروجی مدار به ازای قوانین فازی جدول (۱)

[1] Mel-frequency cepstral coefficients (MFCC)

[2] Zero Crossing Rate (ZCR)

جدول ۲- مقایسه نتایج دقت روش‌های مختلف یادگیری ماشین.

ML methods (%)	ANFIS	ANN	WHO-SNN	WHO-FWS-SNN
Digit recognize	81.2	84.38	92.3	97.2
IRIS	95.85	96.27	95.47	98.93

۵- شبیه‌سازی و بحث نتایج

۵-۱- مجموعه داده‌ها و استخراج ویژگی‌ها

مجموعه داده TIDIGITS کلیپ‌های صوتی ضبط شده، که در آن بلندگوها ارقام را با صدای بلند می‌گویند. یک مجموعه داده ساده صوتی/گفتاری شامل ضبط ارقام گفتاری در فرمت wav با فرکانس ۸ کیلوهرتز است. ضبط‌ها به گونه‌ای بریده شده‌اند که در ابتدا و انتهای آن تقریباً حداقل سکوت را داشته باشند. این مجموعه از مراجع [۲۹و۳۰] برداشت می‌شوند. در کارهای مختلف [۳۱و۳۲] عمدتا از ضرایب کپسترال فرکانس مل برای تمایز ارقام مختلف گفتاری استفاده شده است. برای ضرایب محاسبه شده از پنجره سیگنال صوت در این کار از یک مدل ویژگی‌های آماری استفاده شده است که در کنار ویژگی‌های مختلف آشکارساز عبور از صفر و توان سیگنال‌های صوتی، جمعا ۷۷ ویژگی برای هر سیگنال صوتی استخراج می‌شود. به دلیل بالا بودن حجم پردازش محاسباتی در این مطالعه موردی به کمک الگوریتم گرگ خاکستری تنها هفت ویژگی اصلی شناخته و انتخاب می‌شود. معیار انتخاب ویژگی‌ها هم بر اساس بالاترین میزان پراکندگی بین طبقات و بالاترین میزان تراکم در هر طبقه برای تک تک ویژگی‌ها با تابع واریانس در برنامه MATLAB قابل محاسبه می‌باشد. بنابراین هر ویژگی‌ای که کمترین مقدار خروجی داده‌های تست را از روی سیگنال‌های صوتی برای تابع هدف ایجاد کند، به‌عنوان ویژگی برتر انتخاب می‌گردد. این جستجوی ویژگی‌های برتر با کمک الگوریتم گرگ خاکستری انجام می‌شود. در شکل (۱) چارچوب روند آموزش، نمایش داده شده است.

در این مقاله، برای آموزش شبکه پیشنهادی از الگوریتم اسب وحشی استفاده شده است. بنابراین، روش پیشنهادی FWS-SNN با رویکرد الگوریتم اسب وحشی با دو روش یادگیری ماشین، که شامل روش شبکه عصبی پیشخور و شبکه فازی عصبی تطبیقی[1] است برای کلاس‌بندی سیگنال‌های صوتی برای ارقام ۰-۹ مقایسه شده است. این روش‌ها برای مجموعه داده‌های تست که به تعداد سی عضو که بر اساس داده‌های آموزش ۱۲۰ عضو تحت روش‌های مختلف آموزش داده شده است، بررسی می‌گردد . برای این مطالعه موردی، شبکه تعریف شده یک شبکه سه لایه با تعداد نورون‌های [1 8 15] برای تمام شبکه‌های یادگیری ماشین استفاده شده است. جدول (۲) نتایج مقایسه را برای دو مطالعه مورد نظر نشان می‌دهد. همانطور که نشان داده شده است روش پیشنهادی WHO-FWS-SNN توانسته است بالاترین دقت را در مقایسه سایر روش‌های یادگیری ماشین ایجاد کند.

۶- نتیجه‌گیری

در این مقاله، یک طبقه‌بندی‌کننده عصبی جدید یادگیری متوالی برای شبکه عصبی اسپایکینگ، به نام FWS-SNN، برای مشکلات طبقه‌بندی ارقام از روی سیگنال‌های صوتی ارائه شده است. الگوریتم یادگیری متوالی WHO-FWS-SNN بعد از تعریف تعداد لایه‌ها و نورون-های هر لایه به طور خودکار وزن‌دهی و تعیین آستانه نورون‌های میانی و خروجی مورد نیاز را تعیین می‌کند و وزن سیناپسی را با استفاده از سیستم وزن‌دهی فازی تصادفی با مجموعه داده‌های تست تطبیق می‌دهد. الگوریتم یادگیری از قوانین به‌روزرسانی وزن کم هزینه محاسباتی استفاده می‌کند که فقط به دو پارامتر برای هر نورون نیاز دارد. این روش پیچیدگی آموزش شبکه عصبی مورد مطالعه را کاهش می‌دهد. ما همچنین تعدادی از استراتژی‌ها را برای بهینه‌سازی شبکه عصبی اسپایکینگ برای پیاده‌سازی در سخت‌افزار با محدودیت حافظه و انرژی، از جمله تقریبی در محاسبه دینامیک عصبی و کاهش دقت در ذخیره‌سازی وزن‌های سیناپسی، توصیف می‌کنیم. روش وزن‌دهی فازی توانسته است با کاهش ٪۷۱/۹ % از پارامترهای شبکه برای وزن‌دهی نورون‌ها کمک خوبی به کاهش پیچیدگی آموزش شبکه شبکه عصبی اسپایکینگ پیشنهادی بدهد. به عنوان کار آینده، ما آموزش شبکه‌های مکرر نورون‌های اسپکینگ را برای کاربرد تشخیص گفتار برای کلاسه‌بندی ارقام را بررسی خواهیم کرد تا عملکرد تشخیص را بهبود ببخشیم. برای این منظور در ادامه از یک مدل‌سازی فازی نوع دوم برای تعریف نورون‌ها جهت افزایش دقت می‌توان استفاده نمود.

مراجع

[1] R. P. Lippmann, “Speech recognition by machines and humans,” *Speech Communication*, vol. 22, no. 1, pp. 1–15, Jul. 1997, doi: https://doi.org/10.1016/s0167-6393(97)00021-6.

[2] Y. SUH and H. KIM, “Cepstral Domain Feature Extraction Utilizing Entropic Distance-Based Filterbank,” *IEICE Transactions on Information and Systems*, vol. E93-D, no. 2, pp. 392–394, 2010, doi: https://doi.org/10.1587/transinf.e93.d.392.

[3] L. Deng, "Processing of acoustic signals in a cochlear model incorporating laterally coupled suppressive elements," Neural Networks, vol. 5, pp. 19-34, 1992.

[4] G. Raut, A. Biasizzo, N. Dhakad, N. Gupta, G. Papa, and S. K. Vishvakarma, “Data multiplexed and hardware reused architecture for deep neural network accelerator,” *Neurocomputing*, vol. 486, pp. 147–159, May 2022, doi: https://doi.org/10.1016/j.neucom.2021.11.018.

[5] Wachirawit Ponghiran and K. Roy, “Spiking Neural Networks with Improved Inherent Recurrence Dynamics for Sequential Learning,” *Proceedings of the ... AAAI Conference on Artificial Intelligence*, vol. 36, no. 7, pp.

[1] Adaptive Neuro-Fuzzy Inference System (ANFIS)

[21]“Kaur, S.A.A. (2012) Modified Edge Detection Technique Using Fuzzy Inference System. International Journal of Computer Applications, 44, 9-12. - References - Scientific Research Publishing,” *Scirp.org*, 2016. https://www.scirp.org/reference/referencespapers?referenceid=1746187
[22] Khalid Anindyaguna, Noor Cholis Basjaruddin, and Didin Saefudin, “Overtaking assistant system (OAS) with fuzzy logic method using camera sensor,” Jan. 2016, doi: https://doi.org/10.1109/icimece.2016.7910420.
[23] F. Jabr, “John A. Long - Publications List,” *Publicationslist.org*, vol. 14, no. 6, 2021.
[24] E. Zorarpacı and S. A. Özel, “A hybrid approach of differential evolution and artificial bee colony for feature selection,” *Expert Systems with Applications*, vol. 62, pp. 91–103, Nov. 2016, doi: https://doi.org/10.1016/j.eswa.2016.06.004.
[25] M. H. Ali, S. Kamel, M. H. Hassan, M. Tostado-Véliz, and H. M. Zawbaa, “An improved wild horse optimization algorithm for reliability based optimal DG planning of radial distribution networks,” *Energy Reports*, vol. 8, pp. 582–604, Nov. 2022, doi: https://doi.org/10.1016/j.egyr.2021.12.023.
[26]“Dynamical Systems in Neuroscience,” *MIT Press*, Jun. 18, 2024. https://mitpress.mit.edu/9780262514200/dynamical-systems-in-neuroscience/ (accessed Sep. 24, 2024).
[27] F. Ponulak and A. Kasinski, “Introduction to spiking neural networks: Information processing, learning and applications,” *Acta Neurobiologiae Experimentalis*, vol. 71, no. 4, pp. 409–433, 2011, Available: https://pubmed.ncbi.nlm.nih.gov/22237491/
[28] V. Terrier, “Language Recognition by Cellular Automata,” *Handbook of Natural Computing*, pp. 123–158, 2012, doi: https://doi.org/10.1007/978-3-540-92910-9_4.
[29] https://github.com/Jakobovski/free-spoken-digit-dataset/blob/master.
[30]https://www.kaggle.com/datasets/jackvial/freespokendigitsdataset.
[31] J. Wu, E. Yilmaz, M. Zhang, H. Li, and K. C. Tan, “Deep Spiking Neural Networks for Large Vocabulary Automatic Speech Recognition,” *arXiv.org*, 2019. https://arxiv.org/abs/1911.08373
[32] A. Pitti, Mathias Quoy, C. Lavandier, and Sofiane Boucenna, “Gated spiking neural network using Iterative Free-Energy Optimization and rank-order coding for structure learning in memory sequences (INFERNO GATE),” *Neural Networks*, vol. 121, pp. 242–258, Jan. 2020, doi: https://doi.org/10.1016/j.neunet.2019.09.023.
[33] K. Aizawa, Y. Nakamura, and Shin’ichi Satoh, *Advances in Multimedia Information Processing - PCM 2004*. Springer Science+Business Media, 2005. doi: https://doi.org/10.1007/b104117.
8001–8008, Jun. 2022, doi: https://doi.org/10.1609/aaai.v36i7.20771.
[6] F. I. Arce-McShane, B. J. Sessle, C. F. Ross, and N. G. Hatsopoulos, “Primary sensorimotor cortex exhibits complex dependencies of spike-field coherence on neuronal firing rates, field power, and behavior,” *Journal of Neurophysiology*, vol. 120, no. 1, pp. 226–238, Jul. 2018, doi: https://doi.org/10.1152/jn.00037.2018.
[7] S. Navlakha, Z. Bar-Joseph, and A. L. Barth, “Network Design and the Brain,” *Trends in Cognitive Sciences*, vol. 22, no. 1, pp. 64–78, Jan. 2018, doi: https://doi.org/10.1016/j.tics.2017.09.012.
[8] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, “Neuroscience-Inspired Artificial Intelligence,” *Neuron*, vol. 95, no. 2, pp. 245–258, Jul. 2017, doi: https://doi.org/10.1016/j.neuron.2017.06.011.
[9] G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, and K. Friston, “The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields,” *PLoS Computational Biology*, vol. 4, no. 8, p. e1000092, Aug. 2008, doi: https://doi.org/10.1371/journal.pcbi.1000092.
[10]Y. Bengio and Y. LeCun, "Scaling learning algorithms towards AI," Large-scale kernel machines, vol. 34, pp. 1-41, 2007.
[11] N. Vogt, “Machine learning in neuroscience,” *Nature Methods*, vol. 15, no. 1, pp. 33–33, Jan. 2018, doi: https://doi.org/10.1038/nmeth.4549.
[12] J. Wu, C. Xu, D. Zhou, H. Li, and K. C. Tan, “Progressive Tandem Learning for Pattern Recognition with Deep Spiking Neural Networks,” *arXiv.org*, 2020. https://arxiv.org/abs/2007.01204 (accessed Sep. 24, 2024).
[13] A. K. Mukhopadhyay, M. P. Naligala, D. L. Duggisetty, I. Chakrabarti, and M. Sharad, “Acoustic scene analysis using analog spiking neural network,” *Neuromorphic Computing and Engineering*, vol. 2, no. 4, p. 044003, Oct. 2022, doi: https://doi.org/10.1088/2634-4386/ac90e5.
[14] S. Y. A. Yarga, J. Rouat, and S. Wood, “Efficient Spike Encoding Algorithms for Neuromorphic Speech Recognition,” *Proceedings of the International Conference on Neuromorphic Systems 2022*, Jul. 2022, doi: https://doi.org/10.1145/3546790.3546803.
[15]Juan Pedro Dominguez-Morales *et al.*, “Deep Spiking Neural Network model for time-variant signals classification: a real-time speech recognition approach,” Jul. 2018, doi: https://doi.org/10.1109/ijcnn.2018.8489381.
[16] F. Khatami and M. A. Escabí, “Spiking network optimized for word recognition in noise predicts auditory system hierarchy,” *PLOS Computational Biology*, vol. 16, no. 6, p. e1007558, Jun. 2020, doi: https://doi.org/10.1371/journal.pcbi.1007558.
[17] Z. Pan, Y. Chua, J. Wu, M. Zhang, H. Li, and E. Ambikairajah, “An efficient and perceptually motivated auditory neural encoding and decoding algorithm for spiking neural networks,” *arXiv.org*, 2019. https://arxiv.org/abs/1909.01302 (accessed Sep. 24, 2024).
[18] D. Roy, P. Panda, and K. Roy, “Synthesizing Images from Spatio-Temporal Representations using Spike-based Backpropagation,” *arXiv.org*, 2019. https://arxiv.org/abs/1906.08861 (accessed Sep. 24, 2024).
[19]J. Wu, Y. Chua, M. Zhang, H. Li, and K. C. Tan, “A Spiking Neural Network Framework for Robust Sound Classification,” *Frontiers in Neuroscience*, vol. 12, Nov. 2018, doi: https://doi.org/10.3389/fnins.2018.00836.
[20] J. Wu, E. Yilmaz, M. Zhang, H. Li, and K. C. Tan, “Deep Spiking Neural Networks for Large Vocabulary Automatic Speech Recognition,” *arXiv.org*, 2019. https://arxiv.org/abs/1911.08373 (accessed Sep. 24, 2024).