

الحمد لله الذي
خلقنا من
الحمم



دانشگاه صنعتی شاهرود

دانشکده برق و رباتیک

رساله دکتری مهندسی کنترل

تشخیص و کنترل تحمل‌پذیر خطا برای سیستم‌های دینامیکی
به کمک یادگیری تقویتی

نگارنده: امین نوری

استاد راهنما:

دکتر محمد علی صدرنیا

استاد مشاور:

دکتر محمد باقر نقی‌بی سیستانی

بهمن ۱۳۹۸

تقدیم به

پدر بزرگوارم

که با اعتماد به شانه‌های محکم واستوارش توانستم قدم درراه تحصیل، علم وکمال بگذارم. به یاد خوبی‌هایت، به یاد ایثارگری‌هایت، به یاد آن همه گذشت و فداکاری که درحق من کردی این مجموعه را تقدیم می‌کنم.

مادر مهربانم

گل خوشبوی زندگیم که بی‌منت وجودت را وقف من نمودی به پاس این همه فداکاری من هم چکیده‌ای از تلاش علمی‌ام را به پای خوبی‌هایت تقدیم می‌کنم.

پروردگارا

کمک کن

دریابم که وسعت زندگی فراوی من است، اما چنان عزیز که یک لحظه‌اش تباه نباید کرد
کار کنم تا رسیدن به هدف‌هایم و بدانم که بدست خواهد آمد
و به رویاهایم دست برآورم، با توان، به تصمیم و با ایمان
دریابم که جسارت و امید را فرو نگذارم و تردید را نگذارم که مرا دلسرد کند از انجام آنچه در
دل دارم
به یاد نگه دارم که جهان نیازمند آفتاب لبخندهای هرچه بیشتر است
یادم باشد که سهم خود را ادا کنم.

زحمات و تلاش‌های همه اساتید گرامی را پاس داشته و مراتب تقدیر و تشکر قلبی را تقدیم به
ساحت مقدس این فرزندگان می‌نمایم. از اساتید عزیز **آقای دکتر محمد علی صدرنیا و آقای دکتر
محمد باقر نقی‌بی سیستانی** که در مراحل مختلف پروژه با دقت و حوصله ارائه طریق نموده و
اینجانب را راهنمایی فرمودند خاضعانه سپاسگزاری می‌نمایم. از اساتید بزرگوار **آقای دکتر ناصر پریز،
آقای دکتر محمد مهدی فاتح و آقای دکتر علیرضا الفی** که قبول زحمت نموده و داوری این
رساله را بر عهده گرفتند تشکر می‌نمایم.

کمال تشکر را از **خانم مهندس الناز کلهر** که در انجام این رساله مرا یاری نمودند دارم. از
دانشجویان سابق دوره کارشناسی دانشگاه صنعتی سجاد و فعلی کارشناسی ارشد دانشگاه صنعتی
شاهرود **آقای مهندس محمد رضا مروی‌پور، آقای مهندس مهدی دزیانی، آقای مهندس
مصطفی اسدی‌نژاد، خانم مهندس زهره پاشایی و آقای مهندس علی روحانی بسطامی** و همه
عزیزانی که در این دوره به اینجانب کمک کردند تشکر می‌کنم.

از دوستان و هم دوره‌ای‌های بسیار عزیزم آقایان **حمید بهزاد، سید وحید قوشخانه‌ای و
سید محمد احمدی** که از آشنایی و مصاحبت با آنها درس‌های بسیاری آموختم، سپاسگزاری می
کنم.

در پایان از **آقای یونسین** مسئول تحصیلات تکمیلی دانشکده، **خانم بابامحمدی** مسئول امور
دانش آموختگان تحصیلات تکمیلی و **خانم عرب** نهایت تشکر را دارم.

تعهد نامه

اینجانب امین نوری دانشجو دوره دکتری رشته مهندسی کنترل دانشکده مهندسی برق و رباتیک دانشگاه صنعتی شاهرود نویسنده رساله "تشخیص و کنترل تحمل‌پذیر عیب برای سیستم‌های دینامیکی به کمک یادگیری تقویتی" تحت راهنمایی دکتر محمد علی صدرنیا متعهد می‌شوم.

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام دانشگاه صنعتی شاهرود و یا Shahrood University of Technology به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تاثیر گذار بوده‌اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاقی رعایت شده است.

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

چکیده

این رساله به کنترل تحمل‌پذیر عیب با استفاده از روش‌های یادگیری تقویتی می‌پردازد. کنترل تحمل‌پذیر عیب با استفاده از یادگیری تقویتی از چند جنبه مورد بحث و بررسی قرار می‌گیرد. در ابتدا، به خاصیت رفتاری مقاوم بودن الگوریتم یادگیری تقویتی اشاره می‌شود و اینکه به صورت ذاتی قابلیت تحمل‌پذیری عیب را تا حدودی دارا می‌باشد. در ادامه، با شناسایی سیستم سالم و بدون عیب به کمک شبکه‌های عصبی توابع پایه شعاعی، مانده محاسبه می‌شود. اولین نوآوری این رساله ترکیب الگوریتم مسیرهای شایستگی - که یکی از کارآمدترین روش‌های یادگیری تقویتی می‌باشد - و کنترل‌کننده مد لغزشی، جهت کنترل تحمل‌پذیر عیب می‌باشد. ساختار پیشنهادی با ساختارهای پیشین که مبتنی بر ترکیب شبکه عصبی و کنترل‌کننده مد لغزشی می‌باشد، مقایسه می‌گردد و رفتار بهتر ساختار پیشنهادی مطرح می‌گردد. در دومین نوآوری رساله، ابتدا دامنه و نوع عیب به کمک پنجره زمانی محاسبه می‌گردد و از این داده‌ها به همراه حالت‌های سیستم و خطای دریافتی، کنترل تحمل‌پذیر عیب با استفاده از روش مسیرهای شایستگی انجام می‌شود. در انتها نشان داده می‌شود که این ساختار با وجود سادگی، رفتار بسیار بهتری را در کنترل تحمل‌پذیر عیب دارا می‌باشد. جهت بررسی رفتاری کنترل تحمل‌پذیر عیب با توجه به ساختار پیشنهادی، روش مسیرهای شایستگی با روش‌های پایه یادگیری تفاوت گذرا همانند یادگیری-Q و سارسا مقایسه می‌شود. در آخرین نوآوری رساله، با استفاده از شبکه‌های عصبی مصنوعی پیشرو، سیستم سالم و بدون عیب شناسایی می‌گردد و بر اساس آن مانده محاسبه می‌شود. با استفاده از مانده بدست آمده و الگوریتم یادگیری تقویتی پیوسته و ترکیب آن با الگوریتم گام خودکار، کنترل تحمل‌پذیر عیب صورت می‌گیرد. روش‌های پیشنهادی در این رساله بر روی دو سیستم بیماری دیابت و سرطان ملانوما که نوعی سرطان پوست می‌باشد شبیه‌سازی گردیده است. بی‌نیاز بودن به مدل محیط، یکی از مهمترین مزایای الگوریتم‌های یادگیری تقویتی در کنترل سیستم‌ها می‌باشد. ولی از آنجا که مدل واقعی (بیمار) در دسترس نمی‌باشد جهت بررسی ساختارهای پیشنهادی برای کنترل تحمل‌پذیر عیب، از مدل ریاضی سیستم بیماری دیابت و سرطان ملانوما استفاده می‌شود. در درمان سرطان ملانوما در حضور عیب، نیمه عمر دارو هم لحاظ گردیده است که در مسایلی که هدف تعیین دوز بهینه دارو می‌باشد بسیار مهم می‌باشد منتهی در کارهای بسیار کمی این مورد لحاظ گردیده است.

کلید واژه: کنترل تحمل‌پذیر عیب، یادگیری تقویتی، مسیرهای شایستگی، شبکه‌های عصبی مصنوعی، یادگیری-Q

Journals:

1. **Amin Noori**, Mohammad Ali Sadrnia, Mohammad Bagher Naghibi-Sistani
"Using Sliding Mode Controller and Eligibility Traces for Controlling the Blood Glucose in Diabetic Patients at the Presence of Fault" 2019, *International Journal of Engineering*.
2. **Amin Noori**, Mohammad Ali Sadrnia, Mohammad Bagher Naghibi-Sistani,
"Fault Tolerant Control of Blood Glucose Concentration Using Reinforcement Learning ", 2020, *International Journal of Industrial Electronics, Control and Optimization (IECO)*.
۳. **امین نوری**، محمد علی صدرنیا، الناز کلهر، سارا صبوری راد " کنترل جمعیت سلول‌های سرطانی در مدل غیرخطی سرطان ملانوما با لحاظ عدم قطعیت با استفاده از الگوریتم یادگیری Q تحت سیاست استدلال مبتنی بر مورد (CBR)" نشریه انجمن مهندسين برق و الکترونیک ایران (jiaeee)، ۱۳۹۸.

Conferences:

4. **Amin Noori**, Mohammad Ali Sadrnia, Mohammad Bagher Naghibi-Sistani,
"Glucose level control using Temporal Difference methods." In *2017 Iranian Conference on Electrical Engineering (ICEE)*, pp. 895-900. IEEE, 2017.
5. **Amin Noori**, Mohammad Ali Sadrnia, Mohammad Bagher Naghibi-Sistani,
"Designing model-free control law for nonlinear systems based on reinforcement learning." In *2017 Iranian Conference on Electrical Engineering (ICEE)*, pp. 901-905. IEEE, 2017.
۶. **امین نوری**، محمد علی صدرنیا، محمد باقر نقیبی-سیستانی "کنترل تحمل‌پذیر عیب در سیستم تاخیردار پالمبو با استفاده از یادگیری تقویتی و الگوریتم گام خودکار" کنفرانس مهندسی برق ایران، یزد، ۱۳۹۸

فهرست مطالب

عنوان	صفحه
فصل ۱- مقدمه	۱
۱-۱- مقدمه‌ای بر اهمیت کنترل تحمل پذیر عیب و استفاده از روش‌های هوشمند در کنترل تحمل پذیر عیب	۱
۲-۱- مروری بر کارهای گذشتگان	۴
۳-۱- ساختار رساله	۹
فصل ۲- مقدمه‌ای بر کنترل تحمل پذیر عیب	۱۱
۱-۲- تعریف عیب	۱۱
۲-۲- انواع دسته‌بندی عیب	۱۵
۳-۲- کنترل تحمل پذیر عیب: دیدگاه عمومی	۱۷
۲-۳-۱- تطبیق	۲۱
۲-۳-۲- روش سوئیچینگ	۲۱
۲-۳-۳- پیش‌بینی	۲۲
۲-۳-۴- روش شبه معکوس	۲۲
۴-۲- آشکارسازی، جداسازی و تشخیص عیب	۲۲
۲-۴-۱- FDI بر اساس تولید مانده	۲۳
۲-۴-۲- تخمین عیب	۲۴
۲-۴-۳- FDI بدون نیاز به مدل (FDI هوشمند)	۲۴
۵-۲- افزونگی	۲۴
فصل ۳- یادگیری تقویتی	۲۷
۱-۳- مقدمه	۲۷
۲-۳- معرفی یادگیری تقویتی و اجزای آن	۲۷
۱-۲-۳- تابع پاداش	۲۹
۲-۲-۳- تابع ارزش	۳۰
۳-۲-۳- خط مشی (سیاست)	۳۲
۳-۳- روش‌های حل مساله یادگیری تقویتی	۳۵
۱-۳-۳- برنامه‌ریزی پویا (DP)	۳۵
۲-۳-۳- مونت کارلو (MC)	۳۶
۳-۳-۳- تفاوت گذرا (TD)	۳۷

۴۱	۳-۳-۴- مسیرهای شایستگی (ET).....
۴۳	فصل ۴- مدل ریاضی بیماری دیابت و سرطان ملانوما (سیستم‌های مطالعه موردی).....
۴۳	۱-۴- بیماری دیابت
۴۵	۲-۴- مروری بر اقدامات صورت گرفته جهت کنترل غلظت گلوکز - انسولین بیمار دیابتی
۴۸	۳-۴- مدل‌های ریاضی بیماری دیابت
۴۸	۳-۴-۱- مدل ریاضی دانکومار و سندیا.....
۴۹	۳-۴-۲- مدل ریاضی پالمبو.....
۵۱	۴-۴- سرطان ملانوما.....
۵۱	۴-۵- سرطان پوست
۵۲	۴-۵-۱- انواع سرطان پوست
۵۴	۴-۶- تشخیص سرطان ملانوما با استفاده از روش‌های پزشکی.....
۵۴	۴-۷- مدل ریاضی سرطان ملانوما.....
۵۹	فصل ۵- تشخیص عیب به کمک تولید مانده و کنترل تحمل‌پذیر عیب با ترکیب یادگیری تقویتی و الگوریتم گام خودکار
۵۹	۵-۱- مقدمه
۶۰	۵-۲- تشخیص خطا به کمک تولید مانده با استفاده از شبکه‌های عصبی مصنوعی
۶۱	۵-۲-۱- شبکه عصبی پیش‌رو و الگوریتم پس‌انتشار خطا
۶۳	۵-۳- کنترل تحمل‌پذیر عیب به کمک الگوریتم گام خودکار و یادگیری تقویتی
۶۶	۵-۴- شبیه‌سازی
۶۸	۵-۴-۱- شبیه‌سازی در عدم حضور عیب
۷۰	۵-۴-۲- شبیه‌سازی رفتار کنترل‌کننده پیشنهادی در حضور عیب.....
۷۵	فصل ۶- کنترل تحمل‌پذیر عیب با استفاده از ترکیب مسیرهای شایستگی و کنترل مدلفزشی
۷۵	۶-۱- مقدمه
۷۵	۶-۲- ساختار کلی
۸۱	۶-۳- ساختار پیشنهادی
۸۴	۶-۳-۱- توابع پایه شعاعی
۸۶	۶-۴- شبیه‌سازی ساختارهای کنترلی فوق بر روی مدل دان کومار و سندیا.....
۸۷	۶-۴-۱- کنترل میزان قند خون با استفاده از شبکه‌های عصبی
۸۸	۶-۴-۲- کنترل میزان قند خون با استفاده از مسیرهای شایستگی
۹۲	۶-۴-۳- اعمال عدم قطعیت در پارامترهای سیستم و شرایط اولیه.....
۹۴	۶-۴-۴- اعمال نویز به سیستم در حضور عدم قطعیت در پارامترهای سیستم و شرایط اولیه.....

۹۶.....	۶-۴-۵- اعمال عدم قطعیت در پارامترهای سیستم در طول درمان
۹۸.....	۶-۴-۶- اعمال تغییرات در شرایط اولیه بیمار
۱۰۰.....	۶-۵- شبیه سازی ساختارهای کنترلی فوق بر روی مدل تاخیر دار پالمبو
۱۰۲.....	۶-۵-۱- عیب وارد شده به سنسور
۱۰۳.....	۶-۵-۲- عیب وارد شده به عملگر
۱۰۵.....	۶-۵-۳- عیب وارد شده به سنسور و عملگر
۱۰۹.....	فصل ۷- کنترل تحمل پذیر عیب با استفاده از روش مسیرهای شایستگی و در نظر گرفتن پنجره زمانی
۱۰۹.....	۷-۱- مقدمه
۱۰۹.....	۷-۲- کنترل تحمل پذیر عیب با استفاده از روش مسیرهای شایستگی و در نظر گرفتن پنجره زمانی
۱۱۲.....	۷-۳- شبیه سازی
۱۱۸.....	۷-۴- شبیه سازی روش های یادگیری تقویتی در کنترل سرطان ملانوما و کنترل تحمل پذیر عیب
۱۱۹.....	۷-۴-۱- تعریف پارامترهای یادگیری تقویتی
۱۲۳.....	۷-۴-۲- شبیه سازی روش یادگیری-Q برای کنترل سرطان ملانوما
۱۳۲.....	۷-۴-۳- کنترل جمعیت سلول های سرطانی در حضور عیب سنسور با استفاده از مسیرهای شایستگی
۱۳۵.....	فصل ۸- نتیجه گیری و پیشنهادات
۱۳۸.....	مراجع:

فهرست شکل‌ها

صفحه

عنوان

شکل ۱-۱: بلوک دیاگرام تشخیص عیب در سیستم وسایل نقلیه زیرزمینی [۴۰].....	۷
شکل ۲-۱: ساختار کنترلی کلی پیشنهاد شده در [۴۱].....	۸
شکل ۳-۱: ساختار یادگیری تقویتی عمیق پیشنهاد شده.....	۸
شکل ۱-۲: چند نمونه عیب نوعی در محرک [۲].....	۱۲
شکل ۲-۲: چند نمونه عیب نوعی در سنسور [۱].....	۱۳
شکل ۳-۲: دسته بندی عیب از دیدگاه مدل کردن عیب: عیب جمع شونده و ضرب شونده.....	۱۵
شکل ۴-۲: دسته بندی عیب از دیدگاه مشخصه زمانی.....	۱۶
شکل ۵-۲: نواحی کاری مختلف FTC [۲].....	۱۸
شکل ۶-۲: سیستم کنترل نوعی تحمل پذیر عیب فعال [۴۲].....	۱۹
شکل ۷-۲: دسته بندی FTC [۲].....	۲۰
شکل ۸-۲: دسته بندی FDI.....	۲۳
شکل ۱-۳: مدل استاندارد یادگیری تقویتی.....	۲۸
شکل ۲-۳: زنجیره حالت و عمل.....	۲۸
شکل ۳-۳: انتخاب اعمال توسط سیاست E-حریصانه.....	۳۳
شکل ۴-۳: به روز رسانی ارزش در روش DP.....	۳۶
شکل ۵-۳: به روز رسانی ارزش در روش MC.....	۳۷
شکل ۶-۳: به روز رسانی ارزش در روش TD.....	۳۸
شکل ۷-۳: توالی حالت-عمل.....	۳۸
شکل ۸-۳: نگاه به عقب در روش مسیرهای شایستگی [۵۱].....	۴۱
شکل ۹-۳: نمای کلی از روند به روزرسانی مسیرهای شایستگی.....	۴۲
شکل ۱-۴: کنترل حلقه بسته بیماری دیابت با عملگر تزریق پمپ.....	۴۵
شکل ۲-۴: نمایش سه لایه پوست.....	۵۱
شکل ۳-۴: تصویری از سلول‌های سرطانی بنیادی یا پایه(سلول‌های بارازل) [۸۵].....	۵۲
شکل ۴-۴: تصویری از سلول‌های سرطانی فلسی پوست (سلول‌های سنگ فرشی) [۸۵].....	۵۲
شکل ۵-۴: تصویری از سرطان ملانوما.....	۵۳
شکل ۶-۴: نمایی از نحوه‌ی تشکیل ضایعه ملانوما.....	۵۳
شکل ۱-۵: شناسایی و کنترل خطای سیستم کنترل کننده‌ی بیماری دیابت توسط ترکیب شبکه‌ی عصبی و یادگیری تقویتی.....	۶۰
شکل ۲-۵: نمایش شبکه عصبی پیشرو.....	۶۱
شکل ۳-۵: یک گام از الگوریتم گام خودکار.....	۶۶
شکل ۴-۵: غلظت گلوکز بیمار پس از کنترل توسط کنترل کننده‌ی یادگیری تقویتی.....	۶۹
شکل ۵-۵: غلظت انسولین پلاسمای خون بیمار پس از کنترل.....	۷۰
شکل ۶-۵: تخمین سیستم توسط شبکه‌ی عصبی چندلایه.....	۷۱
شکل ۷-۵: آشکارسازی خطای سیستم با روش شبکه‌ی عصبی.....	۷۲
شکل ۸-۵: میزان تغییرات گلوکز خون در حضور عیب در حضور کنترلر پیشنهادی و در عدم حضور آن.....	۷۳

- شکل ۵-۹: میزان تغییرات انسولین خون در حضور عیب در حضور کنترلر پیشنهادی و در عدم حضور آن..... ۷۳
- شکل ۶-۱: سیستم کنترل تحمل‌پذیر عیب با ترکیب کنترل مدل‌گشتی و شبکه‌های عصبی مصنوعی..... ۷۶
- شکل ۶-۲: کنترل میزان گلوکز با استفاده از روش مسیرهای شایستگی در زمان رخداد عیب..... ۸۲
- شکل ۶-۳: شمای کلی از نحوه به‌روز رسانی ارزش حالت-عمل در روش مسیرهای شایستگی..... ۸۳
- شکل ۶-۴: کنترل میزان گلوکز با استفاده از روش مسیرهای شایستگی..... ۸۴
- شکل ۶-۵: نمایش توابع پایه شعاعی (نمای داخلی شبکه)..... ۸۵
- شکل ۶-۶: عیب وارد شده به عملگر سیستم..... ۸۹
- شکل ۶-۷: عیب وارد شده به سنسور سیستم..... ۸۹
- شکل ۶-۸: میزان گلوکز در یک بیمار مبتلا به دیابت در حضور هر دو عیب سنسور و عملگر..... ۹۰
- شکل ۶-۹: میزان انسولین در یک بیمار مبتلا به دیابت در حضور هر دو عیب سنسور و عملگر..... ۹۰
- شکل ۶-۱۰: میزان دوز داروی تزریقی به بیمار با استفاده از روش یادگیری تقویت و شبکه‌های عصبی در هنگام رخداد عیب..... ۹۱
- شکل ۶-۱۱: میزان گلوکز خون در چهار بیمار دیابتی مختلف با استفاده از مسیرهای شایستگی و شبکه‌های عصبی..... ۹۳
- شکل ۶-۱۲: میزان انسولین خون در چهار بیمار دیابتی مختلف با استفاده از مسیرهای شایستگی و شبکه‌های عصبی..... ۹۳
- شکل ۶-۱۳: میزان میزان دوز داروی تزریقی به چهار بیمار دیابتی مختلف با استفاده از مسیرهای شایستگی و شبکه‌های عصبی..... ۹۴
- شکل ۶-۱۴: گلوکز خون در حضور عیب و نویز در بیمار مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی..... ۹۵
- شکل ۶-۱۵: انسولین خون در حضور عیب و نویز در بیمار مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی..... ۹۵
- شکل ۶-۱۶: دوز داروی تزریقی به بیمار مبتلا به دیابت در حضور عیب و نویز با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی..... ۹۶
- شکل ۶-۱۷: گلوکز خون در حضور عدم قطعیت در پارامترهای سیستم در طول درمان در رخداد عیب در بیمار مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی..... ۹۷
- شکل ۶-۱۸: انسولین خون در حضور عدم قطعیت در پارامترهای سیستم در طول درمان در رخداد عیب در بیمار مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی..... ۹۷
- شکل ۶-۱۹: میزان دوز داروی تزریقی در حضور عدم قطعیت در پارامترهای سیستم در طول درمان در رخداد عیب در بیمار مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی..... ۹۸
- شکل ۶-۲۰: گلوکز خون در حضور تغییر شرایط اولیه و رخداد عیب در بیمار شماره ۱ مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی..... ۹۹
- شکل ۶-۲۱: انسولین خون در حضور تغییر شرایط اولیه و رخداد عیب در بیمار شماره ۱ مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی..... ۹۹
- شکل ۶-۲۲: میزان دوز داروی تزریقی در حضور تغییر شرایط اولیه و رخداد عیب در بیمار شماره ۱ مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی..... ۱۰۰
- شکل ۶-۲۳: عیب وارد شده به عملگر سیستم..... ۱۰۱
- شکل ۶-۲۴: عیب وارد شده به سنسور سیستم..... ۱۰۱
- شکل ۶-۲۵: میزان گلوکز در یک بیمار مبتلا به دیابت در حضور عیب سنسور با استفاده از روش‌های مسیرهای شایستگی، بازیگر-نقاد و شبکه‌های عصبی..... ۱۰۲
- شکل ۶-۲۶: میزان انسولین در یک بیمار مبتلا به دیابت در حضور عیب سنسور با استفاده از روش‌های مسیرهای شایستگی، بازیگر-نقاد و شبکه‌های عصبی..... ۱۰۳

شکل ۶-۲۷: میزان گلوکز در یک بیمار مبتلا به دیابت در حضور عیب وارد شده به عملگر با استفاده از روش‌های مسیرهای شایستگی، بازیگر- نقاد و شبکه‌های عصبی ۱۰۴

شکل ۶-۲۸: میزان انسولین در یک بیمار مبتلا به دیابت در حضور عیب وارد شده به عملگر سیستم با استفاده از روش‌های مسیرهای شایستگی، بازیگر- نقاد و شبکه‌های عصبی ۱۰۴

شکل ۶-۲۹: میزان گلوکز در یک بیمار مبتلا به دیابت در حضور عیب سنسور و عملگر با استفاده از روش‌های مسیرهای شایستگی، بازیگر- نقاد و شبکه‌های عصبی ۱۰۵

شکل ۶-۳۰: میزان انسولین در یک بیمار مبتلا به دیابت در حضور عیب وارد شده به سنسور و عملگر سیستم با استفاده از روش‌های مسیرهای شایستگی، بازیگر- نقاد و شبکه‌های عصبی ۱۰۶

شکل ۷-۱: کنترل میزان گلوکز با استفاده از روش مسیرهای شایستگی در زمان رخداد عیب در سنسور ۱۱۰

شکل ۷-۲: انتخاب اعمال توسط سیاست $\epsilon - greedy$ ۱۱۰

شکل ۷-۳: پنجره زمانی در نظر گرفته شده برای تشخیص عیب ۱۱۱

شکل ۷-۴: نمای کلی شبکه عصبی جهت آموزش برای تشخیص دامنه و نوع عیب ۱۱۲

شکل ۷-۵: کنترل میزان گلوکز با استفاده از ترکیب مسیرهای شایستگی و کنترل مد لغزشی در زمان رخداد عیب در سنسور (حالت بهبود یافته فصل ۶) ۱۱۲

شکل ۷-۶: عیب وارد شده به سنسور سیستم ۱۱۴

شکل ۷-۷: تشخیص دامنه عیب وارد شده به سنسور سیستم توسط روش شبکه‌های عصبی ۱۱۵

شکل ۷-۸: تغییرات میزان گلوکز در حضور عیب وارد شده به سنسور سیستم توسط روش مسیرهای شایستگی و ترکیب روش مسیرهای شایستگی و کنترل مد لغزشی (RL+SMC) ۱۱۵

شکل ۷-۹: میزان انسولین تزریق شده به بیمار توسط روش مسیرهای شایستگی و ترکیب روش مسیرهای شایستگی و کنترل مد لغزشی (RL+SMC) ۱۱۶

شکل ۷-۱۰: میزان دوز داروی تزریق شده به بیمار توسط روش مسیرهای شایستگی و ترکیب روش مسیرهای شایستگی و کنترل مد لغزشی (RL+SMC) ۱۱۷

شکل ۷-۱۱: روند عملکرد پارامترهای یادگیری تقویتی در کنترل جمعیت سلول‌های سرطانی ۱۲۱

شکل ۷-۱۲: نمایش گام‌های زمانی تزریق دارو ۱۲۲

شکل ۷-۱۳: روند کاهش میزان دارو در بدن در طول ۱۰۰ ساعت (اثر نیمه عمر داروی اینترفرون) ۱۲۲

شکل ۷-۱۴: رفتار جمعیت سلول‌های سرطانی در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل بهینه، مقدار ثابت و دوز داروی صفر ۱۲۶

شکل ۷-۱۵: رفتار OT1CTL‌های فعال تزریق شده به نقطه تزریق در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل بهینه، مقدار ثابت و دوز داروی صفر ۱۲۷

شکل ۷-۱۶: رفتار OT1CTL‌های فعال در محفظه پوست، در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل بهینه مقدار ثابت و دوز داروی صفر ۱۲۸

شکل ۷-۱۷: تغییرات آنتی‌بادی CD137 تزریق شده به نقطه تزریق در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل بهینه، مقدار ثابت و دوز داروی صفر ۱۲۸

شکل ۷-۱۸: تغییرات آنتی‌بادی CD137 در محفظه پوست، در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل، مقدار ثابت و دوز داروی صفر ۱۲۹

شکل ۷-۱۹: رفتار CTL‌های ساده، در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل بهینه، مقدار ثابت و دوز داروی صفر ۱۳۰

- شکل ۷-۲۰: تغییرات آنتیژن تومور، در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل بهینه، مقدار ثابت و دوز داروی صفر..... ۱۳۱
- شکل ۷-۲۱: عیب وارد شده به سنسور سیستم..... ۱۳۲
- شکل ۷-۲۲: رفتار جمعیت سلول‌های سرطانی در حالت استفاده از روش مسیرهای شایستگی..... ۱۳۳
- شکل ۷-۲۳: رفتار متغیرهای سیستم در حضور عیب سنسور..... ۱۳۳

فهرست جدول‌ها

صفحه

عنوان

جدول ۱-۴: ثوابت سیستم دانکومار و سندیا [۸۳].....	۴۹
جدول ۲-۴: معرفی پارامترهای مربوط به روابط مدل ریاضی پالمبو تاخیری [۸۴].....	۵۰
جدول ۳-۴: معرفی متغیرهای حالت سیستم.....	۵۶
جدول ۴-۴: تعریف ثوابت سیستم.....	۵۷
جدول ۱-۵: معرفی پارامترهای مربوط به روابط مدل ریاضی پالمبو تاخیری.....	۶۷
جدول ۱-۶: ثوابت سیستم [۸۳].....	۸۷
جدول ۲-۶: مجموع دوز داروی تزریقی با استفاده از روش شبکه‌های عصبی و یادگیری تقویتی.....	۹۱
جدول ۳-۶: ثوابت سیستم پالمبو [۶۰].....	۱۰۱
جدول ۴-۶: مجموع دوز داروی تزریقی با استفاده از روش یادگیری تقویتی و تزریق دوز داروی ثابت.....	۱۰۷
جدول ۱-۷: مجموع دوز داروی تزریقی با استفاده از روش مسیرهای شایستگی و ترکیب روش مسیرهای شایستگی و کنترل مد لغزشی (RL+SMC).....	۱۱۷
جدول ۲-۷: مقادیر در نظر گرفته شده برای پارامترهای سیستم [۹۳].....	۱۱۸
جدول ۳-۷: شرایط اولیه برای یک بیمار مبتلا به ملانوما.....	۱۲۵
جدول ۴-۷: مقایسه مجموع داروهای تزریقی به بیمار در سه حالت ثابت، الگوریتم یادگیری Q، مسیرهای شایستگی و با استفاده از روش کنترل بهینه.....	۱۳۲

فصل ۱- مقدمه

۱-۱- مقدمه‌ای بر اهمیت کنترل تحمل پذیر عیب و استفاده از روش‌های هوشمند

در کنترل تحمل پذیر عیب

با افزایش میزان پیچیدگی در سیستم‌ها می‌بایست امنیت و قابلیت اطمینان آن‌ها نیز کاملاً مورد توجه قرار گیرد. با پیچیده‌تر شدن سیستم‌ها و پیچیدگی ساختار آن‌ها، سیستم‌ها در برابر عیوب آسیب‌پذیری بیشتری خواهند داشت. منظور از عیب انحراف غیرمجاز یک یا چند پارامتر سیستم از شرایط مطلوب می‌باشد. تشخیص اولیه‌ی عیوب سیستم، زمانی‌که سیستم هنوز در یک ناحیه‌ی کنترل شده‌ای مشغول به کار است می‌تواند به پیشگیری از روند وقایع غیرعادی کمک کند و از زیان سیستم بکاهد که در نوبه‌ی خود می‌تواند جلوی فجایع ناگهانی و از کار افتادگی سیستم را بگیرد. بنابراین عیب‌یابی یک موضوع مهم و قابل توجه برای کارورزان صنعتی و محققان علمی می‌باشد. تاریخچه‌ی زیادی برای فرآیند عیب‌یابی، از شیوه‌های تحلیلی گرفته تا هوش مصنوعی و روش‌های آماری وجود دارد. از دید مدل‌سازی روش‌هایی وجود دارند که نیاز به مدل دقیق ریاضی فرآیند، مدل‌های نیمه‌کمی یا مدل‌های کیفی دارند. در طرف دیگر قضیه، روش‌هایی وجود دارد که هیچ شکلی از اطلاعات مدل را در نظر نمی‌گیرند و تنها بر داده‌های فرآیندهای قبلی تکیه می‌کنند. به علاوه با داشتن اطلاعات فرآیند، تکنیک‌های دشواری برای جستجوی عیب وجود دارد که می‌تواند برای عیب‌یابی مورد استفاده قرار گیرد. شیوه‌های عیب‌یابی بحث شده می‌تواند به دو دسته طبقه‌بندی شود، شیوه‌های برگرفته از داده‌ها و شیوه‌های بر اساس مدل فرآیند [۱].

بهینه‌سازی عملکرد سیستم‌های بحرانی همانند هواپیما، سیستم‌های شیمیایی، سیستم‌های مربوط به نیروگاه‌های هسته‌ای، سیستم‌های مربوط به علم پزشکی و ... بسیار مهم می‌باشد. در این سیستم‌ها نیاز می‌باشد تا سیستم‌های کنترلی قابلیت کنترل تغییرات وسیع در شرایط عملکرد سیستم را داشته

باشند. به هر حال بعضی سناریوهای ناخواسته و اتفاقات غیر معمول در سیستم باعث می‌شود که کنترلر عملکرد و پایداری تضمین شده را از دست بدهد. نمونه‌هایی از این اتفاقات ناخواسته شامل عیب‌ها^۱، خرابی^۲ و آسیب دیدن سیستم^۳ می‌باشد که به صورت معمول در فرآیند طراحی کنترلر در نظر گرفته نمی‌شود. مسئله بدست آوردن مرحله‌ای از عملکرد مطلوب و پایداری در زمانی که این اتفاقات ناخواسته رخ می‌دهد نیاز به یک استراتژی متفاوت نسبت به کنترلرهای معمول دارد. این مسئله به خصوص برای سیستم‌های بحرانی و حیاتی (همانند سیستم‌های شیمیایی و پزشکی) و سیستم‌های پر هزینه (همانند ماهواره) بیشتر نمود پیدا می‌کند. به عنوان مثال هواپیما که در آن عیبی در سنسورها و یا محرک‌ها رخ می‌دهد و یا اینکه خود سیستم دچار سانحه می‌شود نیازمند یک استراتژی متفاوت برای کنترل سیستم دارای نقصان می‌باشد [۲].

این موارد باعث مطرح شدن استراتژی کنترلی تحت عنوان کنترل تحمل پذیر عیب^۴ (FTC) شده است. روش‌های کنترلی متفاوتی برای مسئله‌ی FTC تاکنون مطرح شده است که می‌توان از این میان به برنامه‌ریزی بهره^۵ [۳]، کنترل پیش‌بین^۶ [۴-۶]، روش‌های شبه معکوس^۷ [۷]، کنترل مقاوم^۸ [۸-۱۱]، کنترل تطبیقی^۹ [۹، ۱۲، ۱۳]، کنترل مد لغزشی^{۱۰} [۱۴-۱۶]، کنترل بر پایه مفاهیم هوش مصنوعی [۱۷-۲۵] و ... اشاره کرد.

مسئله FTC را می‌توان به دو دسته فعال^{۱۱} و غیرفعال^{۱۲} تقسیم کرد. در ساختار غیر فعال به صورت پایه از خاصیت مقاوم بودن کنترل‌کننده استفاده می‌شود. در این دسته ساختار کنترلی معمولاً دارای

¹ Faults

² Failures

³ System damages

⁴ Fault Tolerant Control (FTC)

⁵ Gain scheduling

⁶ Model Predictive Control (MPC)

⁷ Pseudo-Inverse Method

⁸ Robust Control

⁹ Adaptive Control

¹⁰ Sliding Mode Control

¹¹ Active

¹² Passive

پیچیدگی کمتری می‌باشند ولی برای آنکه برای بدترین حالت رخداد عیب تاثیر لازم را داشته باشند کارایی لازم را ندارند. کنترل‌کننده‌های تحمل‌پذیر عیب فعال فقط در زمانی که عیب رخ می‌دهد، واکنش لازم را انجام می‌دهند. اطلاعات مربوط به رخداد عیب معمولا از بخش دیگری که تحت عنوان "تشخیص عیب و جدا سازی عیب"^۱ شناخته می‌شود، بدست می‌آید. کنترل‌کننده‌های تحمل‌پذیر عیب در حالت فعال خاصیت پیکره‌بندی دوباره را دارا می‌باشند. همانطور که مشخص است کنترل‌کننده‌های تحمل‌پذیر عیب فعال دارای انعطاف بیشتری نسبت به نوع غیر فعال می‌باشند. تحقیقات اولیه در این زمینه بر روی روش‌های تصویری^۲ تمرکز کرده بودند. در روش‌های طرح ریزی اگر عیب مشخصی شناسایی شود، قانون کنترلی مورد نظر از مجموعه کنترلهای از پیش مشخص شده و محاسبه شده انتخاب می‌شود و به صورت برخط کنترل‌کننده تغییر می‌کند. مطالعات بعدی بر روی تطبیق برخط و ترکیب کنترل‌کننده‌ها به صورت آنلاین تمرکز کردند. پیکره‌بندی دوباره ساختار کنترلی معمولا در هنگام رخداد عیب‌های جدی همانند خرابی کامل در سنسورها و یا محرک‌ها نیاز می‌باشد. کنترل تحمل‌پذیر عیب هم اکنون نیز یک مسئله تحقیقاتی باز می‌باشد [۲].

زمانی که عیبی در سیستم رخ می‌دهد مسئله مهم متوجه شدن رخ دادن این عیب و نوع عیب می‌باشد. بعد از آن تصمیم‌گیری در مورد نحوه برخورد با این عیب حایز اهمیت می‌باشد. تشخیص عیب و یافتن محل و سپس انجام عمل مناسب پایه‌ی مبحث کنترل تحمل‌پذیر عیب می‌باشد.

اخیرا بحث تشخیص عیب^۳ در فرایندها از روش‌های متداول قدیمی به سمت روش‌های مبتنی بر هوش مصنوعی پیش می‌رود. علت این امر مزایای غیر قابل انکار و برجسته‌ی روش‌های مبتنی بر هوش مصنوعی نسبت به روش‌های دیگر می‌باشد. از جمله این مزایا می‌توان به مستقل بودن از مدل فرایند، سادگی در محاسبات و روابط بین پارامترها، مقاوم بودن در برابر تغییرات، هزینه‌ی کمتر و سرعت بیش‌تر در فاز اجرا و غیره اشاره نمود [۲۶]. شبکه‌های عصبی، یادگیری تقویتی، الگوریتم

¹ Fault detection and Isolation (FDI)

² Projection

³ Fault

ژنتیک و منطق فازی از جمله مهم‌ترین شیوه‌های مبتنی بر هوش می‌باشند که در عیب‌یابی سیستم‌ها می‌توان از آن‌ها بهره برد. به طور کلی از اوایل دهه‌ی ۱۹۷۰ تاکنون با روش‌های مختلفی به عیب‌یابی سیستم‌های دینامیکی متعددی پرداخته شده است. یکی از جدیدترین روش‌هایی که اخیراً در زمینه کنترل تحمل‌پذیر عیب^۱ در زیر شاخه روش‌های هوشمند مورد استفاده قرار گرفته است، یادگیری تقویتی^۲ می‌باشد که در این رساله قصد بر آن است تا با استفاده از مفاهیم یادگیری تقویتی کنترل تحمل‌پذیر عیب را برای سیستم‌های دینامیکی انجام دهیم. سیستم‌های پزشکی، سیستم‌های بسیار حساسی می‌باشند. بنابراین عملکرد صحیح آنها نیز از اهمیت ویژه‌ای برخوردار است. بروز عیب در این سیستم‌ها باعث خسارات جانی جبران‌ناپذیری خواهد شد. بنابراین تشخیص سریع، دقیق و به‌موقع عیب در سیستم‌های پزشکی از مواردی می‌باشد که در چند ساله اخیر بسیار مورد توجه قرار گرفته است. بدین جهت در این رساله روش‌های پیشنهادی کنترل تحمل‌پذیر عیب مبتنی بر یادگیری تقویتی برای دو سیستم پزشکی ۱- دیابت و ۲- سرطان ملانوما مورد بررسی قرار گرفته است. سیستم‌های پزشکی مورد بحث دارای دستگاه معادلات دیفرانسیل غیرخطی برای بیان مدل ریاضی‌شان می‌باشند. شایان ذکر است که اکثر روش‌های مطرح شده برای کنترل تحمل‌پذیر عیب در سیستم‌های خطی مورد استفاده قرار گرفته‌اند.

۱-۲- مروری بر کارهای گذشتگان

در زمینه تشخیص عیب به کمک یادگیری تقویتی کارهای محدودی انجام شده است که شامل موارد زیر می‌باشد: در مقاله [۲۷]، تشخیص عیب در یک سیستم دارای عدم قطعیت با استفاده از یادگیری تقویتی انجام گرفته است. در این مقاله، با تخمین حالت‌ها و محاسبه عیب بدست آمده با سیستم اصلی، تشخیص و شناسایی عیب انجام گرفته است. در مقاله [۲۸]، با استفاده از روش‌های یادگیری تقویتی و ترکیب آن با منطق فازی، تشخیص و شناسایی عیب انجام گرفته است. فورستر و

¹ Fault Tolerant Control (FTC)

² Reinforcement Learning (RL)

مورفی^۱ (۲۰۱۱) با استفاده از یادگیری تقویتی عیب و محل رخداد عیب در شبکه‌های سنسوری بی‌سیم^۲ (WSN) را تعیین کرده‌اند [۲۹]. پاتن و همکاران از روش مونت کارلو برای طراحی FDI استفاده کرده‌اند [۳۰].

در بحث تشخیص و کنترل عیب موارد زیر مطرح می‌شود: ایستروپانیان^۳ و همکاران به دیدگاه رباست بودن یادگیری تقویتی پرداخته‌اند و بیان کرده‌اند که الگوریتم‌های حل مسئله یادگیری تقویتی به صورت ذاتی تا حدی دارای خاصیت کنترل عیب می‌باشند بر همین اساس آن‌ها از یادگیری تقویتی در بحث پهنای باند مورد نیاز در کاربردهای پزشکی استفاده کرده‌اند [۳۱]. یانسون از مفاهیم کنترل تحمل پذیر عیب برای یادگیری تقویتی چند عامله استفاده کرده است [۳۲].

در مقاله [۳۳]، برای کنترل عیب در سیستم‌های غیرخطی از روش یادگیری تفاوت‌گذرا^۴ در مسئله یادگیری تقویتی استفاده کرده‌اند و برای طراحی مشاهده‌گر و تخمین حالت‌ها نیز، به دلیل وجود نویز و اغتشاش در سیستم از فیلتر کالمن تعمیم‌یافته استفاده شده است. در مقاله [۳۴]، تشخیص و کنترل عیب در سیستم‌های غیرخطی مورد بررسی قرار گرفته است. از دو روش یادگیری تقویتی و شبکه‌های عصبی برای این کار استفاده شده است و در نهایت نشان داده شده است که روش یادگیری تقویتی بهتر و سریع‌تر از شبکه‌های عصبی توانسته است عیوب را تشخیص دهد و سیستم را در حضور آن کنترل کند. در بخش یادگیری تقویتی از روش بازیگر-نقاد استفاده کرده است. در مقاله [۳۵]، برای ردیابی بهتر سیستم خطی و کنترل آن در برابر عیوب از روش محاسبه مانده و ترکیب آن با تئوری H_{∞} برای تشخیص عیب در سیستم استفاده کرده‌اند. از روش یادگیری Q در بخش یادگیری تقویتی برای ردیابی بهینه استفاده کرده‌اند. در این مقاله با تعیین آستانه‌ی تطبیقی برای تشخیص عیب سیستم، مانع از تشخیص اشتباه عیوب و نویزهای سیستم شده‌اند. در روش یادگیری تقویتی از الگوریتم یادگیری Q به عنوان سیاست کنترلی و ردیابی بهینه استفاده شده است. از یک زیرفضای

¹ A. Forster, A. L. Murphy

² Wireless Sensor Network (WSN)

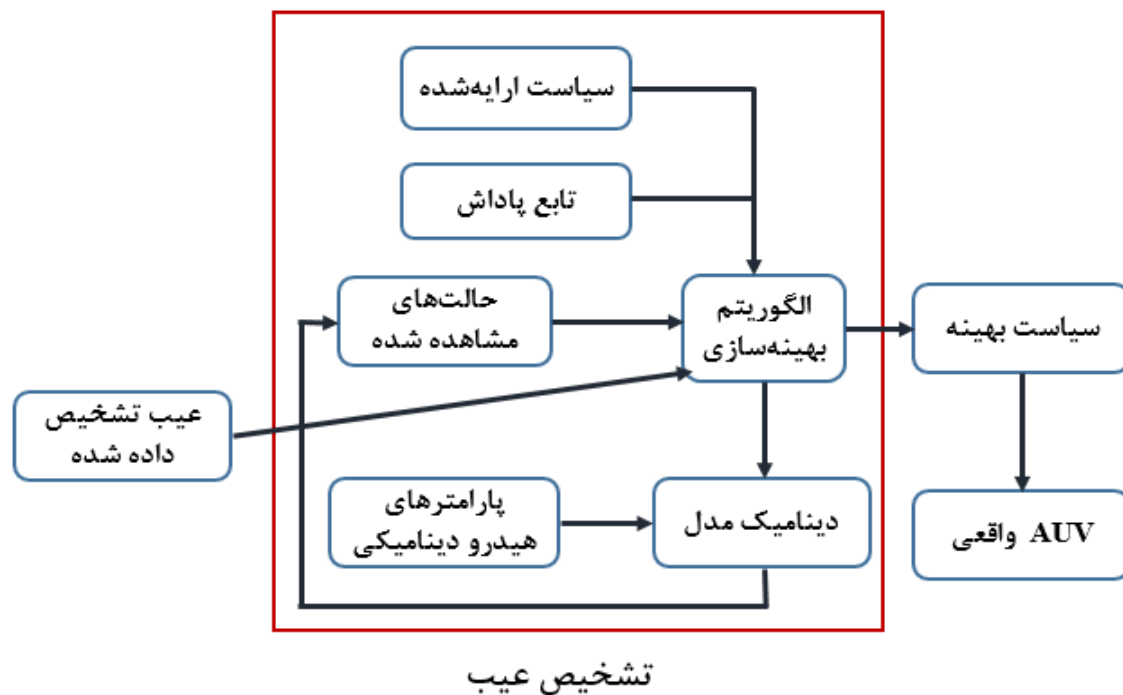
³ Istepanian

⁴ Temporal Difference

کمکی جهت طراحی مشاهده گر بر اساس مانده استفاده شده است. برای کاهش نویز محیط و عملکرد مناسب سیستم از تئوری H_{∞} استفاده شده است. در نهایت نتایج حاصل از شبیه سازی نشان داده است که عیوب را در سیستم های خطی گسسته در زمان ردیابی و کنترل کند. در مقاله [۳۶]، با فرض اینکه هیچ اطلاعاتی از عیب وارد شده به سیستم وجود ندارد، از روش یادگیری تقویتی برای تشخیص و کنترل آن استفاده شده است. در مقاله [۲۰]، مسئله تشخیص و کنترل عیب در سیستم های غیر خطی پیوسته در زمان مطرح شده است. در این مقاله دو سیاست در نظر گرفته شده است: (۱) کنترل بهینه سیستم ها در عدم حضور عیب (۲) کنترل سیستم در حضور عیب. برای تحلیل پایداری سیستم از نظریه ی لیپانوف استفاده شده است و روش استفاده شده برای کنترل عیب در این سیستم ها روش بازیگر- نقاد می باشد که یکی از روش های حل مسئله یادگیری تقویتی می باشد. در مقاله [۲۴]، کنترل و تشخیص عیب در سیستم های چند متغیره گسسته در زمان با استفاده از روش یادگیری تقویتی انجام گرفته است. کنترل سیستم به صورت ردیابی خروجی و ردیابی عیب می باشد و انتخاب اعمال نیز با استفاده از شبکه های عصبی انجام گرفته است. در مقاله [۳۷]، از یک الگوریتم پیشرفته به نام "گام خودکار"^۱ در مسئله یادگیری تقویتی استفاده شده است. در این مقاله، روش پیشنهاد شده با روش تخمین حالت های سیستم با استفاده از روش کمترین مربعات بازگشتی مقایسه شده است و در نهایت نشان داده شده است که استفاده از یادگیری تقویتی و الگوریتم "گام خودکار" با دقت بالاتر و همچنین سرعت همگرایی بیشتر توانسته است عیب را در سیستم تشخیص دهد و آن را از بین ببرد. از دیگر کارهای مشابه در این زمینه می توان به [۳۸، ۳۹] اشاره نمود. در مقاله [۴۰]، تشخیص و کنترل عیب در وسایل نقلیه مورد استفاده در زیر دریا با استفاده از یادگیری تقویتی چندهدفه انجام گرفته است. بدین صورت که با تشخیص و شناسایی عیب در این سیستم ها، سیستم کنترلی ساختار خود را تغییر می دهد. با انتخاب سیاست بهینه، بهترین مسیر را برای رسیدن به هدف و کنترل تحمل پذیر عیب در سیستم انتخاب می کند. نتایج نشان داده است که این روش قادر بوده است که عیوب وارده

¹ Auto Step

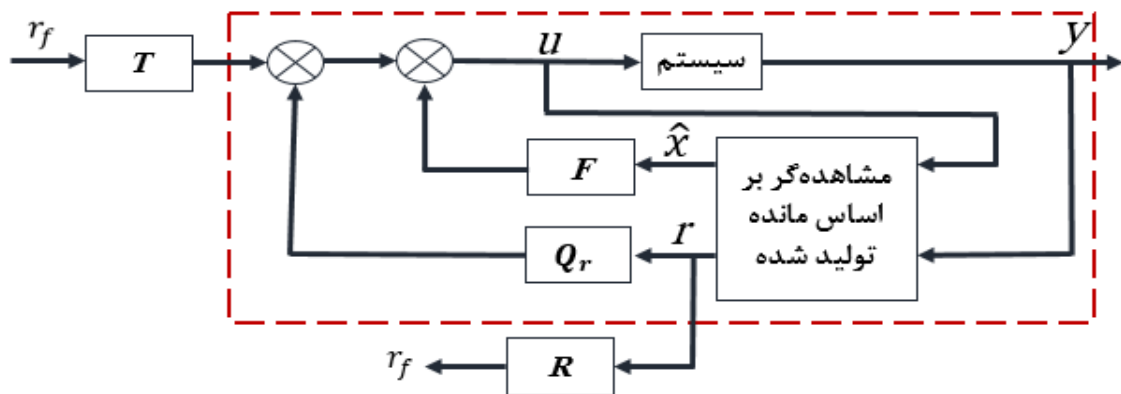
راتشخیص داده و سپس کنترل تحمل‌پذیر عیب را انجام دهد. شکل (۱-۱) بهبود و تشخیص عیب در سیستم‌های وسایل نقلیه زیرزمینی را نشان می‌دهد.



شکل ۱-۱: بلوک دیاگرام تشخیص عیب در سیستم وسایل نقلیه زیرزمینی [۴۰]

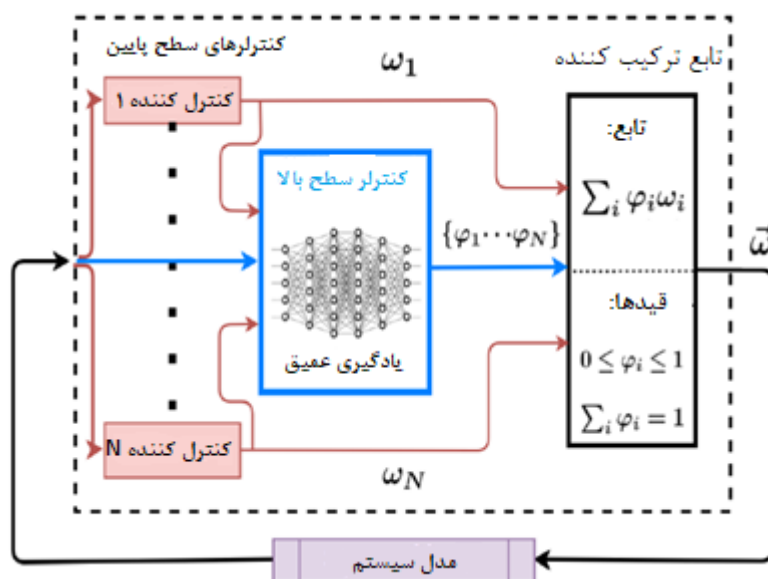
همانطور که در شکل فوق نشان داده شده است، ابتدا عیب وارد شده تشخیص داده می‌شود و سپس بهبود و رفع آن توسط روش یادگیری تقویتی انجام می‌گیرد.

در [۴۱]، از روش الگوریتم یادگیری Q جهت کنترل عیب در سیستم‌های مبتنی بر داده‌های عددی و بدون استفاده از مدل ریاضی سیستم استفاده شده است. طرح پیشنهادی توانسته است در محیط‌های دارای نویز و وقوع اعمال تصادفی و همچنین محیط‌های واقعی و تصادفی به خوبی عمل کرده و کنترل عیب و رفع آن را تسهیل کند. در این مقاله روش پیشنهادی بر روی داده‌های بدست‌آمده از موتور DC اعمال شده است. نتایج نشان داده است که این روش به خوبی توانسته است عیوب را تشخیص داده و کنترل کند. ساختار کنترلی کلی ارائه شده در این مقاله در شکل (۲-۱) نشان داده شده است.



شکل ۲-۱: ساختار کنترلی کلی پیشنهاد شده در [۴۱]

در شکل فوق، فیلتر R برای تنظیم عملکرد تشخیص عیب وارده به سیستم، فیلتر T برای تنظیم عملکرد ردیابی خروجی سیستم می‌باشد. بلوک F نیز پایداری حلقه بسته سیستم را تضمین می‌کند. در [7]، برای تشخیص عیوب ناشناخته در ردیابی مسیر کوادکوپتر از روش یادگیری تقویتی عمیق استفاده شده است. در این مقاله ساختار کنترل کننده ترکیبی از کنترلرهای سطح پایین (کنترل کننده PID، کنترل کننده پیش‌بین)، کنترل کننده سطح بالا (شبکه‌های عصبی) و ترکیب این دو می‌باشد. در شکل (۳-۱) این ترکیب قابل نمایش می‌باشد.



شکل ۳-۱: ساختار یادگیری تقویتی عمیق پیشنهاد شده

همانطور که در شکل فوق نشان داده شده است، خروجی‌های مدل سیستم به عنوان حالت‌های ورودی برای کنترلر بیان شده در نظر گرفته شده است.

۱-۳- ساختار رساله

در فصل دو مقدماتی در مورد تعاریف پایه عیب صحبت خواهد گردید. سپس مفاهیم تشخیص و کنترل تحمل‌پذیر عیب مطرح خواهد شد. در فصل سه مقدماتی در رابطه با یادگیری تقویتی صحبت خواهد شد و مهمترین الگوریتم‌های یادگیری تقویتی شامل الگوریتم‌های تفاوت‌گذرا و مسیرهای شایستگی معرفی خواهند شد. در فصل چهار سیستم‌های مطالعه موردی رساله شامل بیماری دیابت و سرطان ملانوما معرفی خواهند گردید و مدل‌های ریاضی آنها معرفی خواهد گردید. در فصل پنج اولین روش تشخیص و کنترل تحمل‌پذیر عیب با استفاده از یادگیری تقویتی و شبکه‌های عصبی مطرح خواهد شد. و سپس بر روی مدل دیابت شبیه سازی خواهد گردید. در فصل شش کنترل تحمل‌پذیر عیب با ترکیب الگوریتم مسیرهای شایستگی و کنترل مد لغزشی معرفی خواهد شد و بر روی دو مدل از بیماری دیابت شبیه سازی خواهد گردید. در فصل هشت کنترل تحمل‌پذیر عیب با استفاده از یادگیری تقویتی و مفهوم پنجره زمانی بحث خواهد شد و بر روی یک مدل از بیماری دیابت و سرطان ملانوما شبیه‌سازی خواهد گردید. در نهایت نتایج و پیشنهادات این رساله در فصل نه مطرح خواهد شد.

فصل ۲- مقدمه‌ای بر کنترل تحمل‌پذیر عیب

۲-۱- تعریف عیب

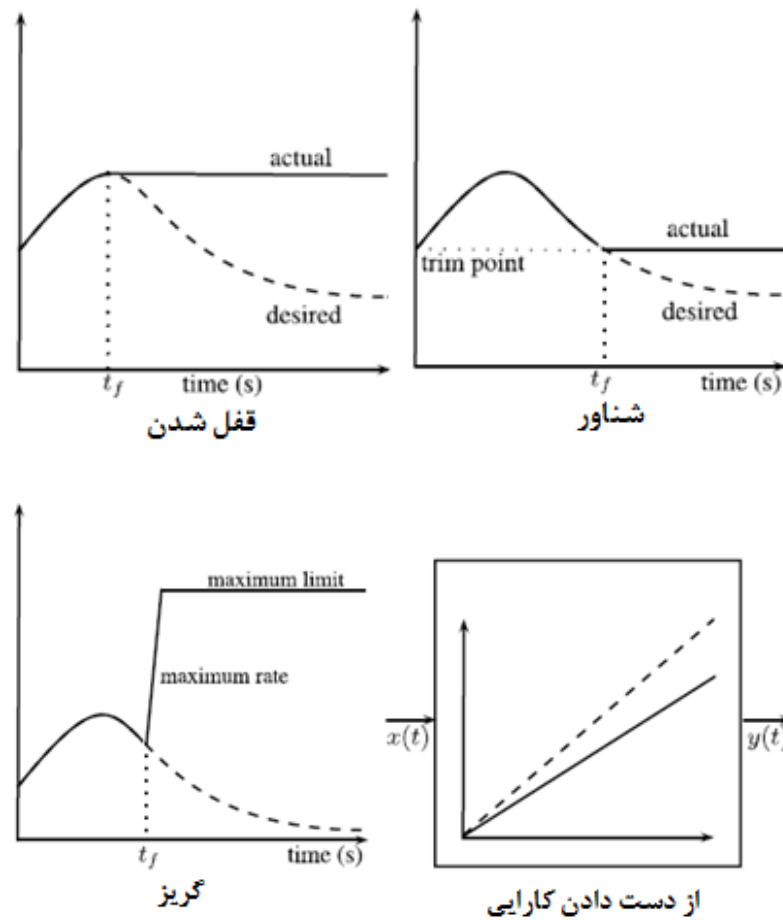
کنترل‌کننده‌های استاندارد (PIDها، کنترل‌کننده‌های پیش‌بین و غیره) به گونه‌ای طراحی می‌شوند که با کم‌رنگ کردن تأثیرات اغتشاش وارده به سیستم، عملکرد سیستم را در شرایط رضایت‌بخشی نگه دارند. گرچه این کنترل‌کننده‌ها می‌توانند، از عهده‌ی انواع مختلفی از اغتشاش برآیند، اما تغییراتی وجود دارند که کنترل‌کننده نمی‌تواند آن‌ها را ساماندهی کند. این تغییرات عیب نامیده می‌شود. به بیان دیگر می‌توان هر گونه انحراف غیرمجاز در حداقل یک رفتار و یا پارامتر مشخص سیستم را عیب تعریف کرد. تعریفی که در اینجا ارایه می‌شود بر اساس تعریف کمیته فنی^۱ IFAC می‌باشد [۱] که به عنوان استاندارد در این زمینه ارایه شده است. بر اساس تعریفی که اینجا برای عیب و خرابی انجام می‌دهیم مفهوم کنترل تحمل‌پذیر عیب نیز ارایه خواهد شد

• **عیب:** انحراف ناخواسته‌ی حداقل یکی از خواص و یا پارامترهای سیستم از شرایط قابل قبول، معمول و یا استاندارد.

• **خرابی:** قطع توانایی یکی از قابلیت‌های سیستم برای انجام عمل مورد نیاز تحت شرایط کاری مشخص

همانطور که از تعاریف فوق مشخص می‌باشد خرابی شرایطی حادث‌تر از رخداد عیب می‌باشد به عنوان مثال زمانی که عیبی در محرک رخ می‌دهد محرک هنوز قابل استفاده می‌باشد ولی ممکن است پاسخ آهسته‌تری داشته باشد و یا کارایی کمتری داشته باشد. ولی وقتی خرابی رخ می‌دهد یک محرک دیگر برای تولید اثر مطلوب مورد نیاز می‌باشد به عنوان مثال در هواپیما عیب‌های مختلفی در محرک داریم که در شکل (۲-۱) برخی از این موارد نشان داده شده است [۲].

¹ International Federation of Automatic Control



شکل ۱-۲: چند نمونه عیب نوعی در محرک [۲]

- **قفل^۱:** در این نوع عیب محرک ثابت می‌شود و قابلیت حرکت ندارد. به عنوان مثال در هواپیما این مورد می‌تواند به خاطر عدم روغن کاری باشد. در پرواز ۱۰۸۰ (L1011 سن دیگو ۱۹۷۷) و پرواز ۹۶ (DC-10 ویندسو، اونتاریو ۱۹۷۲) این عیب رخ داده است.
- **شناور^۲:** در این نوع عیب سطوح کنترلی به صورت آزادانه حرکت می‌کنند بدون اینکه تاثیری بر حرکت هواپیما بگذارند. نمونه‌ای از این نوع عیب می‌تواند به علت از دست دادن مایع هیدرولیک باشد که در سانحه هوایی پرواز ۱۲۳ (بویینگ ۷۴۷ ژاپن ۱۹۸۵) و پرواز DHL-A300b4 (ایرباس ۳۰۰، بغداد ۲۰۰۳) رخ داده است [۲].

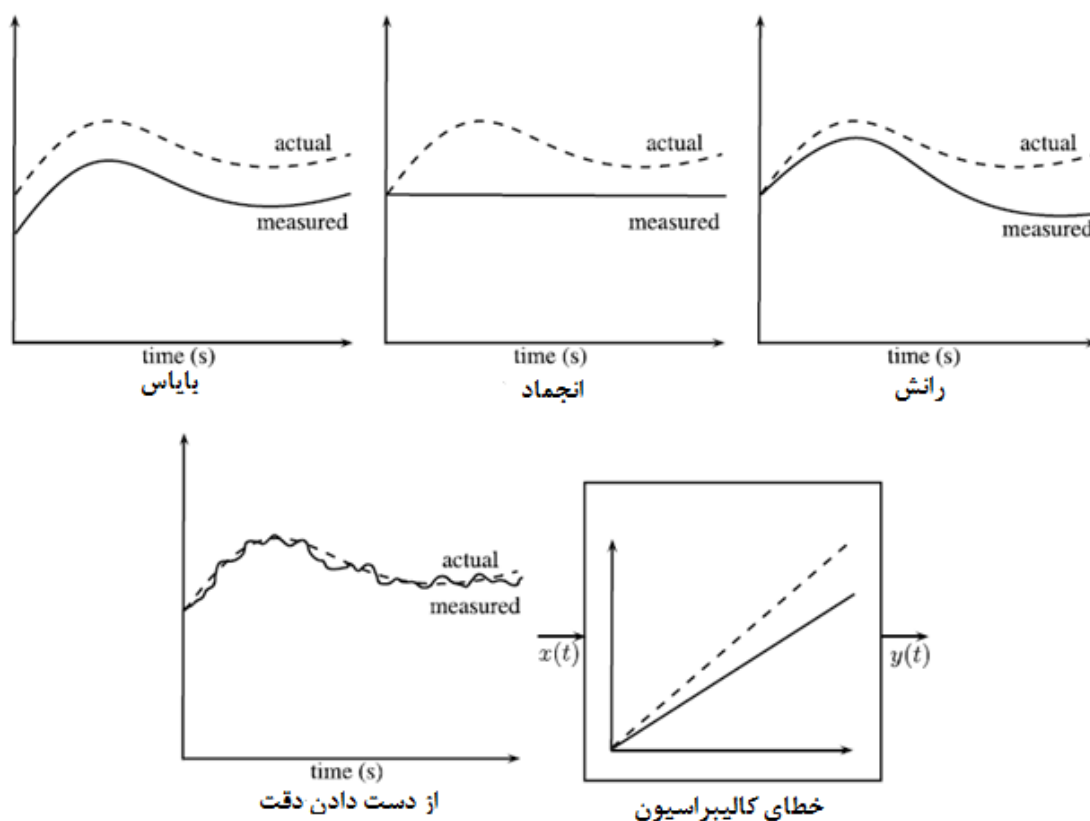
¹ Lock Failure

² Float Failure

- **گریز^۱:** در این مورد سطوح کنترلی با بیشترین نرخ خود حرکت می‌کنند تا اینکه به بالاترین یا پایین‌ترین حد ممکن خود برسند. نمونه این عیب در پرواز ۸۵ (بویینگ ۷۴۷ آنچورج، آلاسکا ۲۰۰۲) رخ داده است [۲].

- **آسیب دیدگی ساختاری^۲:** تا اینجا تمام عیب‌ها مربوط به سطوح کنترلی هواپیما بودند. نوع دیگری از عیب که در هواپیما رخ می‌دهد آسیب دیدگی ساختاری می‌باشد. این آسیب دیدگی می‌تواند باعث تغییر شرایط کاری سیستم شود که مربوط به تغییر در ضرایب آیرودینامیک سیستم می‌باشد [۲] نمونه‌ای از این نوع عیب در پرواز ۹۶۱ (ایرباس ۳۱۰ وورادرو، کوبا ۲۰۰۸) می‌باشد.

شکل (۲-۲) چند نمونه از عیب‌های نوعی در سنسور را نشان می‌دهد



شکل ۲-۲: چند نمونه عیب نوعی در سنسور [۱]

¹ Run away, Hard over

² Structural damage

- **بایاس:** یک عیب ثابت می‌باشد که بین سیگنال اندازه‌گیری شده و سیگنال واقعی می‌باشد.
 - **رانس^۱** شرایطی می‌باشد که در آن خطای اندازه‌گیری با گذشت زمان افزایش می‌یابد.
 - **از دست دادن دقت^۲:** این عیب زمانی رخ می‌دهد که اندازه‌گیری‌ها هیچ موقع مقدارهای صحیح مقدار اندازه‌گیری شده را نشان نمی‌دهند.
 - **انجماد سنسور^۳:** در این حالت سنسور یک مقدار ثابت را به جای مقدار واقعی نشان می‌دهد.
 - **عیب کالیبراسیون^۴:** بیان غلطی از مقدار صحیح متغیر فیزیکی اندازه‌گیری شده از سیگنال الکتریکی می‌باشد.
- عیب‌های سنسور می‌توانند به خاطر خرابی در اجزای تشکیل دهنده سنسور و یا از دست دادن دقت بر اساس استهلاک و کهنگی رخ دهند. نمونه‌ای از عیب رخ داده در سنسور، پرواز ۱۲۴ (بویینگ ۷۷۷ پرث ۲۰۰۵) می‌باشد [۱] که باعث شد کنترل اتوماتیک پرواز تا حدی از دست برود و هواپیما رفتار خشنی از خود نشان دهد که خلبان سیستم خلبان خودکار و سیستم ناوبری را در آن پرواز خاموش کرد.
- می‌توان عیب را بر اساس زمان نیز دسته بندی کرد [۴۲] که از این دیدگاه عیب‌ها به دو دسته‌ی عیب‌های ناگهانی^۵ و عیب‌های ابتدایی^۶ تقسیم می‌شوند
- ۱- عیب‌های ناگهانی: باعث رخ دادن تغییرات ناگهانی و ناخواسته می‌شود که عموماً به راحتی قابل تشخیص می‌باشند. به عنوان مثال گریز نمونه‌ای از این عیب می‌باشد.

¹ Drift

² Loss of accuracy

³ Freezing of sensor

⁴ Calibration error

⁵ Abrupt faults

⁶ Incipient faults

۲- عیب‌های ابتدایی: عیب‌های اولیه‌ای که در سیستم رخ می‌دهند و تغییرات آرامی دارند و به صورت واضح و آشکار دیده نمی‌شوند. به عنوان مثال رانش آرام در سنسور نوعی از این گونه عیب می‌باشد. هرچند این نوع عیب تاثیر آنی زیادی ندارد ولی به هر حال اگر این عیب هم به مدت زمان طولانی نادیده گرفته شود باعث آسیب‌های بعدی می‌شود.

۲-۲- انواع دسته‌بندی عیب

عیب‌ها در حالت کلی به سه دسته زیر تقسیم می‌شوند [۲، ۴۲]:

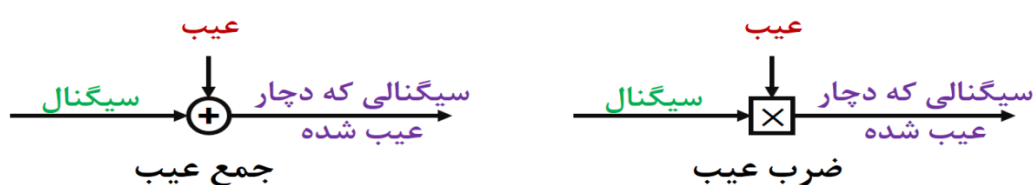
دسته‌بندی‌های مختلفی برای عیب می‌توان ارائه داد. دسته‌بندی می‌تواند براساس مکان رخ‌دادن خطا در سیستم و یا براساس تغییرات زمانی پیشرفت خطا در سیستم باشد. براساس محل خطا می‌توان سه دسته عیب به صورت زیر تعریف کرد.

الف. عیب محرک: که شامل عملکرد نادرست در تجهیزاتی است که سیستم را تحریک می‌کند. به عنوان مثال عیب محرک الکترومکانیکی در یک موتور دیزلی.

ب. عیب فرآیند: هنگامی رخ می‌دهد که تغییرات در سیستم، عدم اعتبار روابط دینامیکی حاکم بر سیستم را به همراه داشته باشد. به عنوان مثال ناشی تانک در یک سیستم کنترل دو-تانکه.

ج. عیب حسگر: که خود را به صورت تغییرات جدی در اندازه‌گیری‌های سیستم نشان می‌دهد.

از دیدگاه دیگر و بسته به مدل کردن عیب در سیستم، عیب‌ها به دو دسته ضرب شونده و جمع شونده تقسیم می‌شوند که این نکته در شکل ۲-۳ نشان داده شده است:



شکل ۲-۳: دسته بندی عیب از دیدگاه مدل کردن عیب: عیب جمع شونده و ضرب شونده

بر اساس مشخصه زمانی نیز می‌توان عیب‌ها را دسته بندی کرد که به سه دسته ناگهانی^۱، نرم^۲ و موقتی^۳ دسته بندی می‌شوند شکل ۲-۴ این عیب‌ها را نشان می‌دهد:



شکل ۲-۴: دسته بندی عیب از دیدگاه مشخصه زمانی

عیبهای ناگهانی اغلب در اثر معیوب شدن سخت افزار به وجود می‌آید این نوع عیب بسیار شدید بوده و بر عملکرد سیستم و پایداری آن تاثیر مستقیم دارد لذا نیاز به اصلاح فوری دارد.

عیبهای نرم بیانگر تغییرات پارامتری کندی بوده که در اثر کهنه شدن سیستم رخ می‌دهد این نوع عیب صدمه ناگهانی ایجاد نمی‌کند ولی شناسایی آن مشکل می‌باشد.

عیبهای موقتی شامل عیبهایی می‌باشد که به صورت مقطعی ظاهر شده و سپس ناپدید می‌گردد.

رخ دادن عیب می‌تواند نتیجه‌ی عوامل متعددی باشد که گاه در مرحله‌ی طراحی و گاه در فاز پیاده سازی به آن‌ها توجه نشده است. یک فرآیند طراحی معمولاً با شرح مساله آغاز می‌شود، سپس طراح براساس درک خود از مساله، جزئیات بیشتری به شرح مساله می‌افزاید. با توجه به شرح مساله یک راه حل سطح بالا ارائه می‌شود و سپس توسعه الگوریتم‌ها و طراحی معماری آغاز می‌گردد. هنگامی که سخت‌افزار و نرم‌افزار تهیه شد، طراحی باید از یک مرحله آزمایشی به صورت مجتمع عبور نماید به طوری که در نهایت عملکرد سیستم چنان باشد که شرح اولیه‌ی مساله را برآورده سازد. مشکلاتی که در هر یک از فازهای طراحی بروز نماید، می‌تواند موجب بروز خطا در سیستم شود. علل خطاهای احتمالی

¹ Abrupt Fault

² Incipient Fault

³ Intermittent Fault

می‌تواند به هر یک از چهار مورد زیر مرتبط باشد: [۱] ۱-توصیف اشتباه ۲-پیاده‌سازی اشتباه ۳-
خرابی عناصر ۴-عوامل خارجی

مورد اول احتمال خرابی به علت اشتباه در توصیف است. این مساله به مواردی از قبیل طراحی الگوریتم اشتباه، معماری سخت افزاری و یا نرم‌افزاری اشتباه برمی‌گردد. مورد بعدی اشتباهات پیاده‌سازی است. پیاده‌سازی دراصل فرآیند تبدیل توصیف سخت‌افزار یا نرم‌افزار به سخت‌افزار واقعی و نرم‌افزار حقیقی است. به علت ضعف پیاده‌سازی گاه ممکن است خطاهایی در این مرحله ایجاد شود. علت دیگر خطا، خرابی عناصر است. ضعف‌های ساخت و خرابی تصادفی هر یک از اجزاء، نمونه‌هایی از این مورد هستند. اجزای فیزیکی به راحتی ممکن است دچار خطا شوند و خرابی هر یک از این عناصر مهم‌ترین علت خطاست. آخرین علت بروز خطا اغتشاشات خارجی محیط هستند. برای مثال تشعشعات و تداخل‌های الکترومغناطیسی یا اشتباهات اپراتوری و در نهایت عوامل محیطی می‌توانند موجب خطا شوند. گاهی وجود عدم قطعیت در محیط و نپرداختن صحیح به این مساله می‌تواند موجب بروز خطا شود. اگر یک مدار الکترونیکی در یک کاربرد نظامی در سرمای بیش از اندازه قرار گیرد، ممکن است نتایج غیرقابل قبول تولید نماید

۲-۳- کنترل تحمل پذیر عیب: دیدگاه عمومی

همانطور که بیان شد بسیاری از تحقیقات در زمینه کنترل تحمل پذیر عیب شامل حل مشکلات رخ داده در سیستم‌های بحرانی، حیاتی و امنیتی همانند هواپیما می‌باشد. البته به کاربردهای دیگری در زمینه‌های دیگر از جمله روبات، سیستم‌های فضایی، سیستم‌های زیر دریایی [۲]، شبکه‌های سنسوری خصوصاً در بحث پزشکی و غیره نیز اشاره شده است.

پاتن در [۴۳] بیان می‌کند که

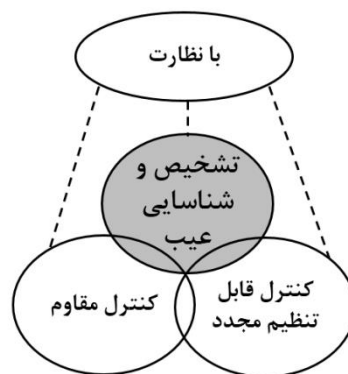
✓ ... تحقیقات مربوط به کنترل تحمل پذیر عیب بسیار مربوط به مشکلات رخ داده در طراحی سیستم هواپیما می باشد. هدف فراهم کردن قابلیت تعمیر خودکار می باشد تا خلبان بتواند هواپیما را به سلامت در هنگام رخداد عیبی جدی به زمین بنشانند.

ژنگ و جیانگ تعریف می کنند [۲]:

✓ ... سیستم های کنترل تحمل پذیر عیب به عنوان سیستم های کنترلی قابلیت حفظ پایداری عام و عملکرد قابل قبول در حالت رخداد عیب و یا خرابی در سیستم را دارا می باشند.

سیستم های کنترل تحمل پذیر عیب همچنین به عنوان خود تعمیر، تجدید ساختار پذیر^۱ و یا سیستم های کنترلی خود طراحی شونده^۲ نیز یاد می شوند.

FTC در واقع ترکیب پیچیده ای از سه شاخه ی تحقیقاتی FDI، کنترل روباست و کنترل تجدید ساختار پذیر می باشد [۴۳].



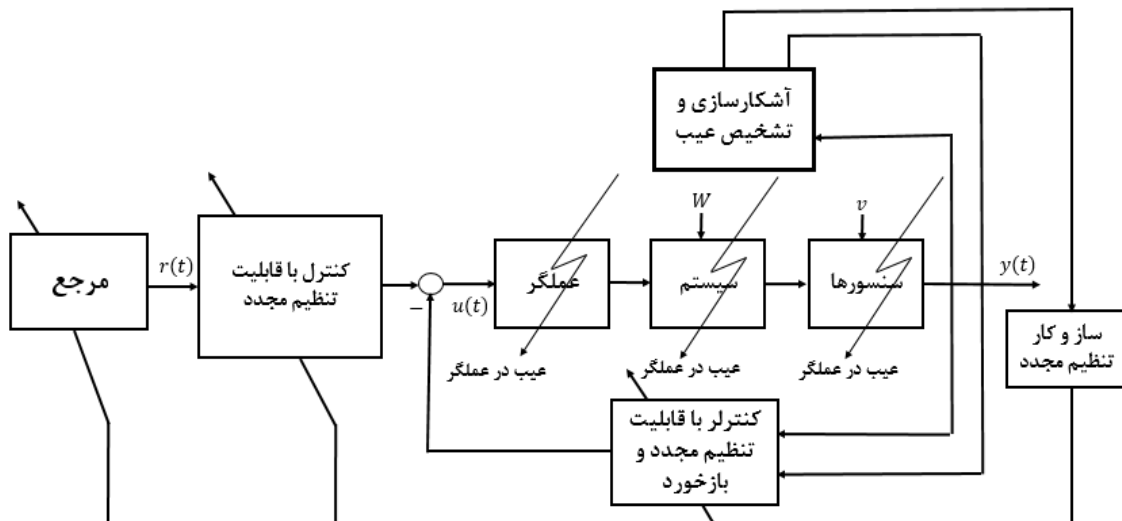
شکل ۲-۵: نواحی کاری مختلف FTC [۲]

پاتن [۴۳] در مورد رابطه ی بین این زمینه های تحقیقاتی صحبت کرده است. در شکل (۲-۶) یک سیستم کنترل نوعی تحمل پذیر عیب فعال^۳ نشان داده شده است.

¹ Restructurable

² Self-designing control systems

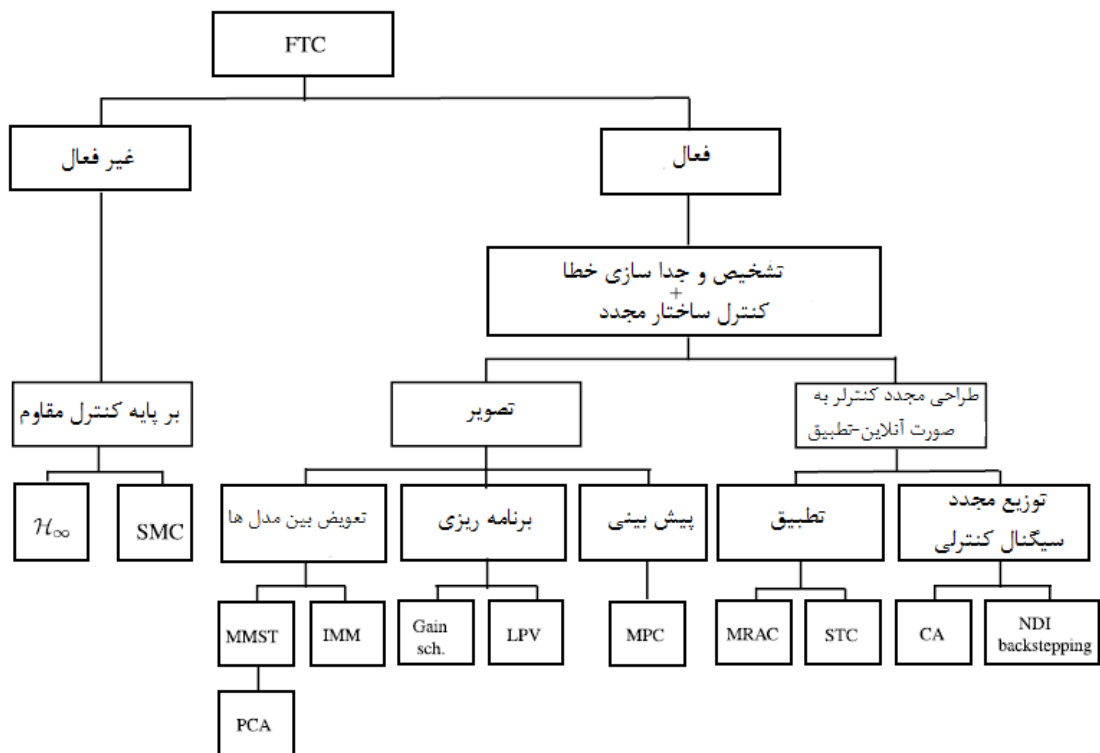
³ Active fault tolerant control



شکل ۶-۲: سیستم کنترل نوعی تحمل پذیر عیب فعال [۴۲]

در بسیاری از ساختارهای FTC زمانی که عیب و یا خرابی در سنسور و یا محرک رخ می‌دهد، قسمت FDI عیب را شناسایی می‌کند و محل عیب را مشخص می‌کند. سپس این اطلاعات به مکانیزم پیکره‌بندی دوباره می‌رسد. سپس کنترلر تجدید پذیر تلاش می‌کند تا خود را به گونه‌ای با عیبی رخ داده تطبیق دهد به گونه‌ای که پایداری و عملکرد سیستم تا حدی حفظ شود. هم FDI و هم کنترلر دارای قابلیت پیکره‌بندی دوباره می‌بایست در برابر عدم قطعیت‌ها و اغتشاشات رباست باشند.

ژنگ و جیانگ [۴۴] و پاتن [۴۳] FTC را به دو دسته مهم تقسیم کرده‌اند که در شکل (۷-۲) نشان داده شده است.



شکل ۲-۷: دسته بندی FTC [۲]

در سیستم‌های کنترل تحمل پذیر عیب پسیو^۱ (PFTCS) کنترل کننده به گونه‌ای طراحی می‌شود تا در برابر عیب و عدم قطعیت مقاوم باشد. بنابراین وقتی که عیبی رخ می‌دهد کنترل کننده طراحی شده از قبل می‌بایست توانایی حفظ پایداری و حفظ نسبی عملکرد مطلوب را داشته باشد. PFTCS نیازی به استفاده از FDI ندارد و همچنین نیاز به کنترل کننده ساختار مجدد و تطبیق ندارد. از طرف دیگر در سیستم‌های کنترل تحمل پذیر عیب فعال^۲ (AFTCS) به عیب‌های رخ داده شده در سیستم به صورت فعال پاسخ داده می‌شود به گونه‌ای که پایداری و عملکرد قابل قبول کل سیستم باقی بماند [۴۴]. بنابراین اکثر AFTCS ها نیازمند FDI جهت تامین اطلاعات مربوط به عیب و خرابی‌ها می‌باشند تا بتوانند پیکره‌بندی دوباره کنترل کننده را انجام دهند. زمانی که عیبی در سیستم رخ می‌دهد.

همانطور که بیان شد در FTC غیر فعال دیدگاه‌های کنترل مقاوم مورد استفاده قرار می‌گیرد و نیاز به FDI نمی‌باشد در حالیکه در FTC فعال اطلاعاتی راجع به رخداد عیب مورد نیاز می‌باشد. AFTC نیز

¹ Passive Fault Tolerant Control Systems

² Active Fault Tolerant Control Systems

خود به دو دسته تقسیم می‌شود: تصویر و کنترل ساختار مجدد و تطبیق. در حالت تصویر کنترل‌کننده‌ها از قبل برای تمام عیب‌ها و یا خرابی‌های ممکن سیستم طراحی می‌شوند. کنترل‌کننده تصویر فقط زمانی فعال می‌شود که عیبی متناظر رخ دهد. تصویر خود به زیر شاخه‌ی گزینش مدل^۱، زمان بندی^۲ و پیش بینی تقسیم می‌شود. شاخه‌ی دیگر AFTC که بر اساس کنترل ساختار مجدد و تطبیق می‌باشد نیز خود به ۲ زیر شاخه‌ی تطبیق و بازسازی سیگنال کنترلی تقسیم می‌شود.

در ادامه به اختصار برخی از روش‌های کنترل ساختار مجدد مورد بررسی قرار می‌گیرند.

۲-۳-۱- تطبیق

کنترل تطبیقی برای اولین بار در سال ۱۹۵۰ برای سیستم اتوپایلت هواپیما طراحی شد. این روش دارای دو رویکرد مستقیم و غیر مستقیم می‌باشد. در رویکرد مستقیم تخمین کنترل‌کننده به صورت مستقیم بدون تخمین پارامترهای سیستم انجام می‌شود. روش غیر مستقیم شامل دو مرحله است، تخمین پارامترهای سیستم و طراحی کنترل‌کننده بر اساس سیستم دو روش رایج در کنترل تطبیقی روش‌های MRAC و STR می‌باشند [۲].

۲-۳-۲- روش سوئیچینگ

ایده استفاده از چندین مدل برای اولین بار در سال ۱۹۹۰ مطرح شد. با توجه به خطی بودن بیشتر روش‌های کنترل کلاسیک، استفاده از طرح مدل‌های چندگانه یک روش ایده آل برای حل مشکل ناشی از تغییرات نقطه کار می‌باشد. در این روش بایستی بانکی از مدل‌ها را داشته باشیم بگونه‌ای که همه عیب‌ها و عدم قطعیت‌های ممکن را شامل شود و برای هرمدل یک کنترل‌کننده طراحی شود. در این روش واحد FDI وظیفه تشخیص نوع عیب را داشته و بر مبنای آن کنترل‌کننده مربوطه انتخاب می‌شود [۲].

¹ Model switching

² scheduling

۲-۳-۳- پیش‌بینی

برعکس خیلی از روش‌های کنترلی که اولین بار توسط جوامع آکادمیک ارائه شدند، روش MPC برای اولین بار توسط صنعت ارائه شد. به همین علت این روش دارای مفاهیم ریاضی ساده‌ای بوده و توسط مهندسين قابل فهم می‌باشد. MPC پرکاربردترین روش در کنترل فرآیندها در صنعت می‌باشد. ایده اصلی روش اینست که به فرآیند اجازه داده شود تا حد ممکن بدون تجاوز به مرزها به محدودیت‌های فرایند اعم از فیزیکی یا امنیتی نزدیک شود تا بتوان تولید را ماکزیمم کرد. این روش یک الگوریتم تکراری براساس کنترل بهینه است [۲].

۲-۳-۴- روش شبه معکوس

ایده روش شبه معکوس طراحی کنترل‌کننده‌ای است که قطب‌های سیستم مربوط به حالت عیب تا حد ممکن بتوانند به قطب‌های حلقه بسته نامی نزدیک باشد [۲]

۲-۴- آشکارسازی، جداسازی و تشخیص عیب

در حالت FTC فعال FDI نقش حیاتی و مهمی را در رابطه با فراهم کردن اطلاعات عیب‌ها و خرابی‌ها در سیستم جهت کنترل ساختار مجدد ایفا می‌کند. مهمترین نقش FDI تشخیص عیب و محل آن می‌باشد کمیته فنی IFAC تعاریف زیر را ارائه داده است [۲]:

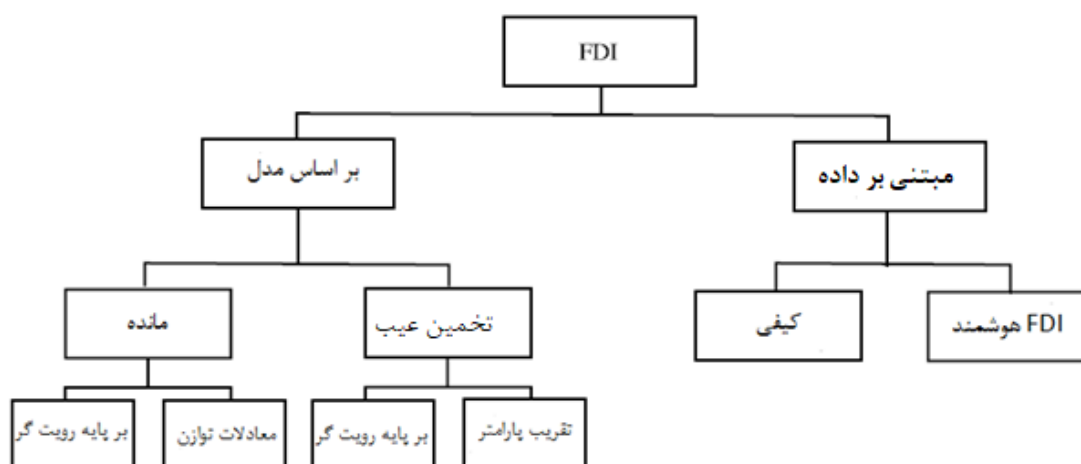
- آشکارسازی عیب^۱: تشخیص عیب‌های حاضر در سیستم و زمان تشخیص
- جداسازی عیب^۲: تشخیص نوع، مکان و زمان آشکار شدن عیب
- تشخیص عیب^۳: شناسایی و تشخیص اندازه و تغییرات عیب

¹ Fault detection

² Fault isolation

³ Fault identification

رابطه‌ی بین FDI و FTC در [۱، ۲، ۴۴] بحث شده است. برای بسیاری از سیستم‌های AFTC، مقاوم بودن FDI تاثیر بسیار زیادی بر مقاوم بودن FTC دارد. دسته‌بندی‌های مختلفی برای FDI وجود دارد. یک دسته‌بندی واضح و آشکار بر اساس فراهم بودن و یا نبودن مدل سیستم می‌باشد. هر کدام از این شاخه‌ها خود به دو دسته قابلیت تقسیم بندی دارند. در حالت FDI بر پایه‌ی مدل دو زیر شاخه اصلی عبارتند از FDI بر اساس مانده و FDI بر اساس تخمین عیب شکل زیر این دسته بندی را نشان می‌دهد.



شکل ۲-۸: دسته بندی FDI

در ادامه چند روش تشخیص عیب به صورت مختصر توضیح داده می‌شود:

۲-۴-۱- FDI بر اساس تولید مانده

در FDI بر اساس تولید مانده، سیگنال‌هایی که از مدل ریاضی و اندازه‌گیری سخت افزاری به دست می‌آیند با یکدیگر مقایسه می‌شوند [۲، ۴۵]. در شرایط رخداد عیب مانده می‌بایست صفر باشد و در زمان رخداد عیب و یا خرابی مانده غیر صفر می‌باشد. سیگنال مانده معمولاً همراه با یک حد آستانه‌ای در نظر گرفته می‌شود تا با اغتشاشات و عدم قطعیت‌ها اشتباه نشود. وقتی که سیگنال مانده از حد آستانه تجاوز کند می‌گویند عیب رخ داده است. معمولاً در تولید مانده عیب و مکان آن تشخیص داده

می‌شود و اطلاعات دیگری در مورد عیب وجود ندارد. سیستم‌های FDI بر اساس مانده قابلیت فهم آسانی دارند و در بسیاری از روش‌ها به راحتی قابل استفاده می‌باشند.

۲-۴-۲ تخمین عیب

بعضی از روش‌های FTC نیازمند این نوع تخمین عیب می‌باشند. یکی از بهترین روش‌ها برای تخمین، فیلتر کالمن می‌باشد.

۲-۴-۳ FDI بدون نیاز به مدل (FDI هوشمند)

یکی از موارد مهم در رابطه با طراحی FDI بر اساس مدل، در دسترس بودن و کیفیت مدل می‌باشد. عیب‌های ناشی از مدل‌های ناکامل یا غیر دقیق بر ساختار عملکرد تشخیص عیب تاثیر گذار می‌باشند [۴۶، ۲]. از اوایل سال ۲۰۰۰ تحقیقات بر روی FDI بدون نیاز به مدل بیشتر شده‌اند بیشتر این تحقیقات بر پایه‌ی هوش مصنوعی و محاسبات نرم می‌باشد همانند شبکه‌های عصبی و منطق فازی [۴۶-۴۹]. یکی از مزایای اسفاده از رویکرد هوشمند برای FDI توانایی مدل کردن هر تابع غیر خطی می‌باشد [۵۰]، همچنین آموزش از طریق نمونه‌ها نیاز به اطلاعات پیشین یا ندارد و یا بسیار کم دارد.

۲-۵-۲ افزونگی^۱

افزونگی را می‌توان به دو دسته تقسیم کرد: مستقیم و تحلیلی، در روش مستقیم، یک نسخه مجدد از ساختار سخت افزاری موجود می‌باشد. به عنوان مثال برای سنسور، ۲ یا ۳ سنسور که یک متغیر را اندازه‌گیری می‌کنند، افزونگی مستقیم می‌باشد. در فرآیندهای معمولی فقط یک سنسور کافی می‌باشد ولی در هنگام رخداد عیب با قرار دادن دو یا سه سنسور اندازه‌گیری با قابلیت اطمینان بالاتری انجام

¹ Redundancy

می‌شود. سیستم رای‌گیری یک نمونه نوعی برای تصمیم‌گیری اینکه کدام کانال‌ها درست عمل می‌کنند و کدام کانال‌ها همراه با عیب هستند می‌تواند مورد استفاده قرار گیرد [۲, ۴۳].

در حالت تحلیلی به جای داشتن چندین سنسور (و یا سخت افزار) یک رویت‌گر که تقریب سیگنال‌های مورد نظر را فراهم می‌کند به عنوان افزونگی مورد استفاده قرار می‌گیرد، در اینحالت هیچ سخت افزار اضافه‌ای مورد استفاده قرار نمی‌گیرد و به جای آن الگوریتمی یا مدل ریاضی یا رویت‌گری در کامپیوتر کنترلی اجرا می‌شود. حالت تحلیلی در بسیاری از سیستم‌هایی که نیاز به کاهش وزن و هزینه دارند مطلوب می‌باشد.

فصل ۳- یادگیری تقویتی

۳-۱- مقدمه

یادگیری تقویتی در یک بیان کلی یعنی با توجه به شناختی که از محیط داریم و بر اساس نتایج تعاملاتی که با محیط داشته‌ایم و سودها و زیان‌هایی که در نتیجه انجام عمل‌های مختلف به دست آورده‌ایم، استراتژی را پیدا کنیم که با عمل به آن در بلند مدت مطلوبیت خود را بیشینه کنیم. الگوریتم‌های یادگیری در حالت کلی به سه دسته "یادگیری با سرپرست"، "یادگیری بدون پرست" و "یادگیری تقویتی" تقسیم می‌شوند. یادگیری تقویتی مابین یادگیری با سرپرست و بدون سرپرست قرار دارد. در روش یادگیری تقویتی عامل یادگیرنده با استفاده از سیگنال‌های پاداش و تنبیه‌ای که از تعامل با محیط به دست می‌آورد، استراتژی خود را بهینه می‌کند. در این بخش در ابتدا به صورت خیلی خلاصه مفاهیم پایه یادگیری تقویتی شرح داده می‌شود. سپس بر روی مفاهیم به‌روزتر که امروزه در یادگیری تقویتی به آن توجه می‌شود تاکید بیشتری می‌شود.

۳-۲- معرفی یادگیری تقویتی و اجزای آن

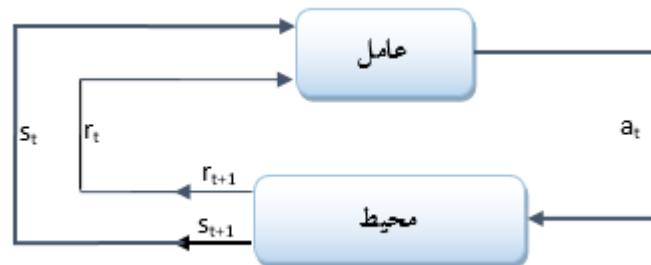
یادگیری تقویتی را می‌توان به صورت ساده در یک جمله بیان کرد: یادگیری بوسیله تعامل با محیط به منظور رسیدن به هدفی مشخص. در حل مسئله با استفاده از روش یادگیری تقویتی دو مورد که نقش بسیار مهمی را ایفا می‌کنند، عامل^۱ و محیط^۲ می‌باشند. عامل با سعی و خطای فراوان در محیط، اطلاعات جدید را هر لحظه از محیط دریافت کرده و خود را به‌روز می‌کند. مسئله یادگیری تقویتی

¹ Agent

² Environment

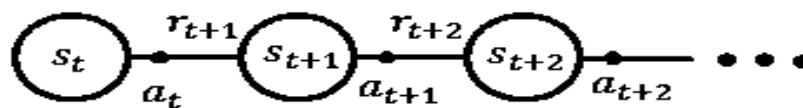
علاوه بر عامل و محیط دارای سه بخش دیگر نیز می‌باشد که شامل عمل^۱، حالت^۲ و پاداش^۳ است.

شکل ۱-۳ تعامل عامل و محیط را نشان می‌دهد.



شکل ۱-۳: مدل استاندارد یادگیری تقویتی

عامل در هر لحظه حالت محیط (s_t) را دریافت می‌کند و با انجام عمل a_t به حالت بعدی (s_{t+1}) می‌رود، در این لحظه پاداش (r_{t+1}) را دریافت می‌کند زنجیره حالت و عمل در شکل ۲-۳ نمایش داده شده است [۵۱].



شکل ۲-۳: زنجیره حالت و عمل

هدف نهایی در مساله یادگیری تقویتی این می‌باشد که مجموع پاداش دریافتی توسط عامل در دراز مدت بیشینه گردد. برای آنکه عامل بتواند برای هر هر حالت، بهترین عمل را انتخاب کند نیاز به سعی و خطا در عمل‌های انجامی در محیط را دارد که نوع جستجو و سعی و خطا با استفاده از خط مشی^۴ یا سیاست مشخص می‌شود که در بخش بعد توضیح داده خواهد شد. [۵۱، ۵۲]

¹ Action

² State

³ Reward

⁴ Policy

۳-۲-۱- تابع پاداش

در یادگیری تقویتی، هدف عامل در قالب سیگنال پاداشی که از محیط دریافت می‌کند بیان می‌شود. در هر مرحله زمانی، این پاداش بصورت عددی ساده بیان می‌شود $r_t \in R$. در بیانی ساده، هدف عامل بیشینه کردن مجموع این پاداش‌هاست. دقت شود بیشینه کردن پاداش در بلندمدت مدنظر است و این لزوماً به معنای بیشینه کردن پاداش در هر مرحله نیست. یک تابع پاداش، نگاشتی از حالت‌های ادراک شده محیط به یک عدد واحد که پاداش نامیده می‌شود، است. یک عامل در طول زمانی که تجربه کسب می‌کند همواره سعی می‌کند که مجموع پاداش دریافتی را بیشتر کند. یک تابع پاداش مشخص می‌کند که چه عملی برای عامل مناسب است و چه عملی مناسب نیست. تابع پاداش را می‌توان با احساس درد و رضایت در سیستم‌های بیولوژیکی مقایسه کرد. در حالت کلی تابع پاداش می‌تواند استوکستیک باشد.

در الگوریتم یادگیری تقویتی عامل بعد از انجام عمل، پاداش خود را که موسوم به پاداش فوری است دریافت می‌کند و حالت بعدی محیط نیز تعیین می‌شود. پاداش دریافتی در این مرحله جدید تحت تاثیر حالت کنونی محیط است بنابراین عامل جهت تصمیم‌گیری علاوه بر تخمین پاداش فوری باید حالت بعد را نیز در نظر بگیرد. در نتیجه مقدار پاداش در آینده باید به گونه‌ای در تصمیم‌گیری در زمان حاضر در نظر گرفته شود. به عبارتی یک عامل باید یاد بگیرد که برای رسیدن به پاداش دلخواه در زمان آینده، عمل مناسب را بیابد و فقط به پاداش در زمان حال حساسیت نشان ندهد.

در این زمینه سه مدل بیشتر مورد توجه قرار گرفته است:

- **مدل افق زمانی محدود:** این مدل ساده‌ترین مدل است. عامل باید پاداش مورد انتظار در h

گام بعدی را در یک زمان داده شده بیشینه سازد. که در معادله (۳-۱) نشان داده شده است.

$$E(\sum_{t=0}^h r_t) \quad (3-1)$$

در رابطه استفاده شده r_t نشان دهنده پاداش عددی آنی دریافت شده در h گام بعدی است. E عملگر امید ریاضی می باشد. در این حالت عامل تنها به لحظه آن را مدنظر قرا می دهد و نگران اتفاقاتی که در آینده می افتد نیست مدل افق زمانی محدود همیشه مناسب نیست، برای مثال در مواردی که مقدار دقیق طول عمر عامل مشخص نیست روش مناسبی به حساب نمی آید.

- **مدل افق زمانی نا محدود با نرخ کاهشی:** در این مدل، عامل میزان پاداش دریافتی در بلند مدت را در نظر می گیرد، ولی در این مدل مقدار پاداش هایی که عامل در بلند مدت می گیرد به صورت هندسی با ضریب کاهش فراموشی، کاهش می یابند که در معادله (۲-۳) نشان داده شده است.

$$E(\sum_{t=0}^h \gamma^t r_t) \quad (3-2)$$

در معادله (۲-۳) γ نرخ فراموشی می باشد که دارای عددی مابین صفر و یک است.

- **مدل میانگین گیری پاداش:** در این مدل عامل سعی می کند اعمالی را انتخاب کند که میزان میانگین پاداش دریافتی را در بلند مدت بیشینه سازد. این مدل مشابه حالت قبل است اگر در آن فرض کنیم که فاکتور فراموشی به سمت یک میل می کند از آنجایی که انتخاب مدل بهینگی و پارامترها اهمیت زیادی دارد، بسیار مهم است که در هر کاری این موارد به خوبی انتخاب شوند. برای اینکه کیفیت هر مدل مورد بررسی قرار می گیرد می توان بعضی از فاکتورها را در نظر گرفت که از جمله این فاکتورها می توان همگرایی نهایی الگوریتم به جواب بهینه، سرعت همگرایی و غیره را نام برد.

۳-۲-۲- تابع ارزش

اگر یک تابع پاداش نشان دهنده خوب یا بد بودن در همان لحظه است یک تابع ارزش مشخص کننده خوب یا بد بودن در بلند مدت است. به صورت کلی، ارزش یک حالت، مجموع کلی پاداش است که

یک عامل انتظار دارد هنگامی که از این حالت حرکت را شروع می‌کند به دست بیاورد. برای مثال، یک حالت ممکن است همیشه پاداش لحظه‌ای پایین داشته باشد اما دارای ارزش بالایی باشد، زیرا از این حالت به حالت‌هایی خواهیم رفت که باعث افزایش پاداش در بلند مدت خواهد شد. عکس این مطلب نیز می‌تواند درست باشد. انتخاب اعمال بر اساس ارزش‌ها صورت می‌گیرد. در یادگیری تقویتی به دنبال اعمالی هستیم که عامل را به حالت‌های با ارزش‌تر با پاداش بالاتر ببرد. متأسفانه مشخص کردن ارزش حالت‌ها کار مشکلی می‌باشد، زیرا پاداش به صورت مستقیم و لحظه‌ای از طرف محیط به عامل داده می‌شود، اما ارزش‌ها باید براساس دنباله‌ای که یک عامل از آن حالت شروع می‌کند. تخمین زده شود. در حقیقت، مهمترین قسمت تمامی الگوریتم‌های یادگیری تقویتی روش تخمین ارزش حالت‌هاست خوشبختانه هسته مرکزی تخمین ارزش مهمترین چیزی است که در طول دهه اخیر در مورد یادگیری تقویتی به خوبی شکل گرفته شده است و در ادامه به آن پرداخته می‌شود.

ارزش یک حالت مشخص s تحت سیاست π که با $V^\pi(s)$ نشان داده می‌شود به صورت معادله (۳-۳) تعریف می‌شود [۵۱]:

$$V^\pi(s) = E_\pi\{R_t | s_t = s\} = E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s\} \quad (3-3)$$

که در فرمول فوق $E_\pi\{\}$ مشخص کننده امید ریاضی می‌باشد و t گام زمانی می‌باشد. V^π را تحت عنوان تابع ارزش حالت^۱ برای سیاست π می‌شناسیم.

به طور مشابه می‌توانیم ارزش عمل a در حالت s تحت استراتژی π را به صورت معادله (۴-۳) تعریف کنیم:

$$Q^\pi(s, a) = E_\pi\{R_t | s_t = s, a_t = a\} = E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a\} \quad (4-3)$$

Q^π را تحت عنوان تابع ارزش حالت-عمل^۲ برای سیاست π می‌شناسیم

¹ State value Function

² Action-State value function

یک خاصیت مهم در توابع ارزش که در یادگیری تقویتی استفاده می‌شود خاصیت رابطه‌ی بازگشتی می‌باشد. برای هر سیاست π و حالت s رابطه‌ی زیر بین ارزش حالت s و حالت‌های بعدی وجود دارد:

$$\begin{aligned}
 V^\pi(s) &= E_\pi\{R_t | s_t = s\} \\
 &= E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s\} \\
 &= E_\pi\{r_{t+1} + \gamma \sum_{k=0}^{\infty} \gamma^k r_{t+k+2} | s_t = s\} \quad (3-5) \\
 &= \sum_a \pi(s, a) \sum_{s'} P_{ss'}^a [R_{ss'}^a + \gamma E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+2} | s_{t+1} = s'\}] \\
 &= \sum_a \pi(s, a) \sum_{s'} P_{ss'}^a [R_{ss'}^a + \gamma V^\pi(s')]
 \end{aligned}$$

معادله (3-5) که به معادله بلمن مشهور است بیان کننده ارزش یک حالت به ارزش حالت‌های بعدی می‌باشد. در معادله (3-5)، $\pi(s, a)$ نشان دهنده سیاست انتخاب عمل در حالت s ، $P_{ss'}^a$ احتمال انتقال از حالت s به حالت s' با انجام عمل a و $R_{ss'}^a$ میزان مجموع پاداش دریافتی (بازگشت^۱) با انجام عمل a در حالت s و رفتن به حالت s' می‌باشد. $V^\pi(s')$ ارزش حالت s' تحت سیاست π را نشان می‌دهد.

۳-۲-۳ خط مشی (سیاست)

یک خط مشی یا سیاست که با (π) نامگذاری می‌شود، شیوه و رفتار عامل یادگیرنده در یک حالت خاص از محیط را مشخص می‌کند. در واقع یک خط مشی، نگاهی از حالتی که عامل در آن قرار دارد به عملی که باید در این حالت انجام دهد است. در خیلی از موارد یک خط مشی ممکن است یک تابع ساده و یا یک جدول جستجو^۲ باشد در حالی که در بعضی از موارد خط مشی می‌تواند یک تابع پیچیده باشد که برای محاسبه آن نیاز به محاسبات زیادی باشد. به بیان ساده‌تر π تابع احتمالی است که احتمال انتخاب شدن هر عمل را در هر حالت و با توجه به گام زمانی می‌دهد. بطور مثال

¹ Return

² Lookup table

$\pi_t(s, a) = p$ می‌گوید که اگر عامل در زمان t در حالت s قرار گرفته باشد، با احتمال p عمل a را انتخاب می‌کند.

از آن جایی که خط مشی به تنهایی برای مشخص کردن رفتار عامل کافی است از خط مشی به هسته یا مغز عامل یاد می‌شود. در حالت کلی خط مشی‌ها می‌توانند به صورت استوکستیک باشند. سیاست‌های مرسوم می‌که در مساله یادگیری تقویتی استفاده می‌شوند، عبارتند از ϵ -حریصانه^۱ و سافت مکس^۲ که در قسمت بعد توضیحات مربوط به این روش‌ها آورده خواهد شد. در این رساله علاوه بر دو سیاست فوق از سیاست جدیدی به نام سیاست استدلال مبتنی بر مورد^۳ نیز استفاده شده است که توضیحات آن نیز در ادامه آورده خواهد شد.

۳-۲-۳-۱ روش ϵ -حریصانه

در این روش انتخاب اعمال طبق رابطه (۳-۶) انجام می‌گیرد.

$$a_t = \begin{cases} a_t^* & \text{with probability } 1 - \epsilon \\ \text{random action with probability } \epsilon \end{cases} \quad (۳-۶)$$

طبق رابطه (۳-۶)، بهترین عمل (a_t^*) با احتمال $1 - \epsilon$ و بقیه اعمال با احتمال یکسان انتخاب می‌شوند [۵۳]. در شکل (۳-۳) روش انتخاب اعمال توسط این سیاست قابل نمایش می‌باشد.

عمل	a_1	a_2	a_3^*	a_4
ارزش عمل	0.01	0.2	1.3	0.4
احتمال انتخاب عمل	$\frac{\epsilon}{n-1}$	$\frac{\epsilon}{n-1}$	$1 - \epsilon$	$\frac{\epsilon}{n-1}$

شکل ۳-۳: انتخاب اعمال توسط سیاست ϵ -حریصانه

^۱ ϵ -greedy

^۲ Softmax

^۳ Case Based Reasoning (CBR)

با توجه به شکل (۳-۳)، عمل a_3 بهترین عمل با بالاترین ارزش بوده است، پس با احتمال $1 - \varepsilon$ انتخاب می‌شود و بقیه اعمال با احتمال یکسان و مقدار $\frac{\varepsilon}{n-1}$ انتخاب می‌شوند که n تعداد اعمال قابل انجام می‌باشد.

۳-۳-۲-۳ سافت مکس

در این روش احتمال انتخاب هر عمل جداگانه و توسط رابطه (۷-۳) انجام می‌گیرد.

$$p(a_t) = \frac{e^{Q_t(a_t)/\tau}}{\sum_{b=1}^n e^{Q_t(b)/\tau}} \quad (۷-۳)$$

τ در رابطه (۷-۳)، پارامتر دما می‌باشد. این پارامتر در اوایل یادگیری برای اینکه احتمال انتخاب تمامی اعمال یکسان باشد و جستجوی در محیط بیشتر شود، مقداری بزرگ انتخاب می‌شود و در گام‌های جلوتر مقدار آن کوچک می‌گردد.

۳-۳-۲-۳ سیاست استدلال مبتنی بر مورد (CBR)

سیاست استدلال مبتنی بر مورد با نام اختصاری CBR یک سیاست اکتشافی شتاب داده شده می‌باشد که برای انتخاب اعمال در الگوریتم‌های یادگیری از آن استفاده می‌شود. بر اساس این سیاست، اگر یک مسئله حل شود و حل آن به هر نتیجه‌ای ختم شود به منظور حل مسئله مشابه در آینده نگه داشته می‌شود که اگر این حل با شکست روبه‌رو شود به منظور جلوگیری از اشتباه در آینده به خاطر سپرده شود. مزیت انتخاب اعمال توسط این سیاست، افزایش سرعت یادگیری و کاهش زمان برای رسیدن به سیاست بهینه و دریافت ماکزیمم مقدار پاداش می‌باشد [۵۴]. در این سیاست از یک تابع اکتشافی با نام $H(s, a)$ برای انتخاب اعمال استفاده می‌شود که در رابطه (۸-۳) نمایش داده شده است.

$$H(s, a) = \begin{cases} \max_i \hat{Q}(s, i) - \hat{Q}(s, a) + \eta & \text{if } a = \pi^H(S) \\ 0 & \text{otherwise} \end{cases} \quad (۸-۳)$$

در این رابطه η یک مقدار حقیقی کوچک و معمولاً یک می‌باشد و باعث وزن‌دهی به عملی که ارزش

بیشتر دارد می‌شود. $\hat{Q}(s, i)$ ارزش حالت و عمل در گام i ام می‌باشد. $\pi^H(s)$ سیاستی است که بر اساس آن انتخاب عمل صورت گرفته است که با استفاده از رابطه (۹-۳) محاسبه می‌شود [۵۴].

$$\pi^H(s) = \begin{cases} \arg \max_a [\hat{Q}(s, a) + \xi H(s, a)^\beta] & \text{if } q < p \\ a_{\text{random}} & \text{otherwise} \end{cases} \quad (۳-۹)$$

در رابطه (۹-۳)، β و ξ پارامترهای کنترلی هستند که مقدار آن‌ها یک در نظر گرفته شده است. q یک عدد تصادفی بین صفر و یک می‌باشد و با استفاده از توزیع نرمال بدست می‌آید. p عددی برای توازن بین جستجو و بهره‌برداری می‌باشد و بین صفر و یک انتخاب می‌شود که هر چه بزرگتر و به یک نزدیک‌تر باشد میزان بهره‌برداری بیشتر خواهد بود و در غیر این صورت هر چه p کوچکتر باشد بهره‌برداری کمتر و جستجو بیشتر خواهد بود. با توجه به رابطه (۹-۳)، اگر q کوچکتر از p باشد، اکشن انتخابی بر اساس اولین معادله در رابطه (۹-۳) محاسبه خواهد شد و در غیر این صورت یک مقدار تصادفی انتخاب می‌شود [۵۴].

۳-۳- روش‌های حل مساله یادگیری تقویتی

برای حل مساله یادگیری تقویتی سه روش پایه شامل برنامه‌ریزی پویا^۱ (DP)، مونت کارلو^۲ (MC) و یادگیری تفاوت گذرا^۳ (TD) می‌باشد. اما در این رساله از روش‌های کارآمدتری تحت عنوان مسیرهای شایستگی نیز بهره برده‌ایم که در ادامه توضیح داده خواهد شد.

۳-۳-۱- برنامه‌ریزی پویا (DP)

در روش برنامه‌ریزی پویا (DP) برای به‌روزرسانی ارزش هر حالت می‌بایست اطلاعات کاملی از احتمال انتقالات به حالت بعدی داشته باشیم به بیان دیگر می‌بایست مدل ریاضی سیستم به صورت کامل در دسترس باشد ولی از آنجایی که یکی از مزایای روش یادگیری تقویتی بی نیاز بودن به مدل ریاضی

^۱ Dynamic Programing (DP)

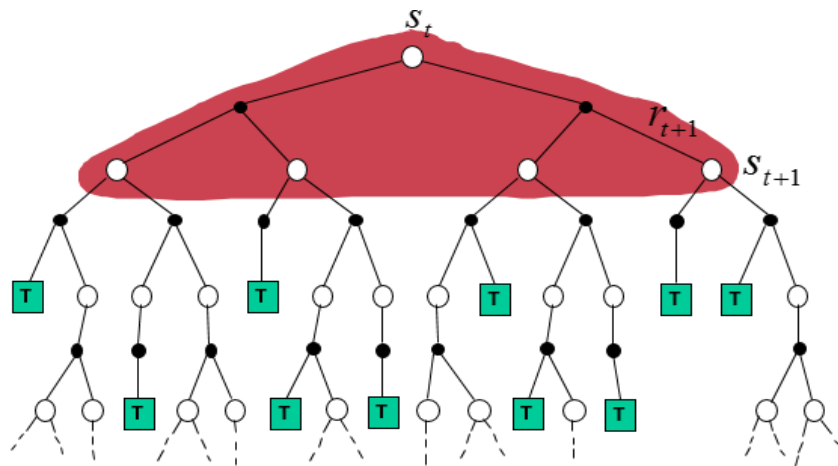
^۲ Monte Carlo (MC)

^۳ Temporal difference (TD)

محیط می‌باشد بسیاری از مراجع این روش را در دسته روش‌های حل مساله یادگیری تقویتی دسته‌بندی نمی‌کنند. معادله به‌روزرسانی ارزش هر حالت در معادله (۳-۱۰) نشان داده شده است.

$$V(s_t) = E_{\pi}\{r_{t+1} + V(s_{t+1})\} \quad (3-10)$$

شکل (۳-۴) نیز شمای کلی به‌روزرسانی ارزش حالت s را نشان می‌دهد.



شکل ۳-۴: به‌روزرسانی ارزش در روش DP

در شکل (۳-۴) دایره‌های تو خالی نشان دهنده حالت‌ها می‌باشند و دایره‌های توپر نشان دهنده اعمال می‌باشند. خانه‌های مربعی شکل که با حرف T نمایش داده شده‌اند نشان دهنده پایان هر اپیزود می‌باشند. همانطور که نشان داده شده است برای به‌روزرسانی ارزش حالت s نیاز به دانستن احتمال انتخاب هر یک از دو عمل مربوطه (دایره‌های تو پر) را داریم و علاوه بر این می‌بایست احتمال اینکه با انجام هر عمل به چه حالتی می‌رویم را نیز داشته باشیم که در بسیاری از مسایل واقعی این اطلاعات در دسترس نمی‌باشد.

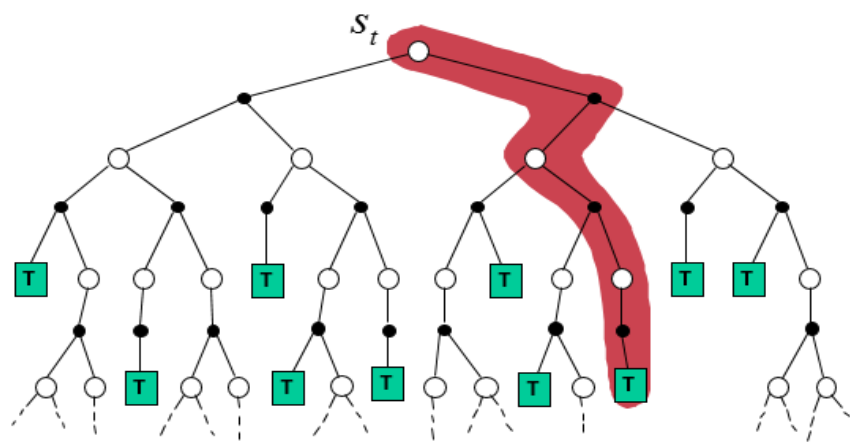
۳-۳-۲ مونت کارلو (MC)

در این روش نیازی به دانستن اطلاعات محیط نداریم و با سعی و خطایی که در محیط انجام می‌دهیم ارزش حالت‌ها را به‌روز می‌کنیم. بنابراین با شروع از هر حالت اولیه و انجام عمل‌ها به پایان یک اپیزود

می‌رسیم که در این مسیر می‌توان ارزش تمام حالت‌ها یا ارزش حالت-عمل‌ها را به‌روز کنیم. معادله (۱۱-۳) به‌روز رسانی ارزش هر حالت را نشان می‌دهد.

$$V(s_t) = V(s_t) + \alpha[R_t - V(s_t)] \quad (۳-۱۱)$$

که در رابطه (۱۱-۳) α نرخ یادگیری می‌باشد. شکل (۵-۳) نیز نشان دهنده شمای کلی به‌روز رسانی ارزش حالت s می‌باشد.



شکل ۵-۳: به‌روز رسانی ارزش در روش MC

۳-۳-۳ تفاوت گذرا (TD)

اگر بخواهیم از بخشی به عنوان هسته یادگیری تقویتی یاد کنیم قطعاً این بخش یادگیری تفاوت گذرا^۱ (TD) می‌باشد [۵۱]. یادگیری تفاوت گذرا مزیت‌های دو روش پایه دیگر یادگیری تقویتی، برنامه‌ریزی پویا^۲ (DP) و مونت کارلو^۳ (MC) را دارا می‌باشد این دو خاصیت عبارتند از:

- یادگیری بدون نیاز به داشتن مدل محیط. یادگیری بر اساس تجربیات خام
- به‌روز رسانی توابع ارزش آن‌ا انجام می‌شود و تا پایان اپیزود منتظر نمی‌ماند.

معادله (۱۲-۳) به‌روز رسانی ارزش هر حالت را نشان می‌دهد.

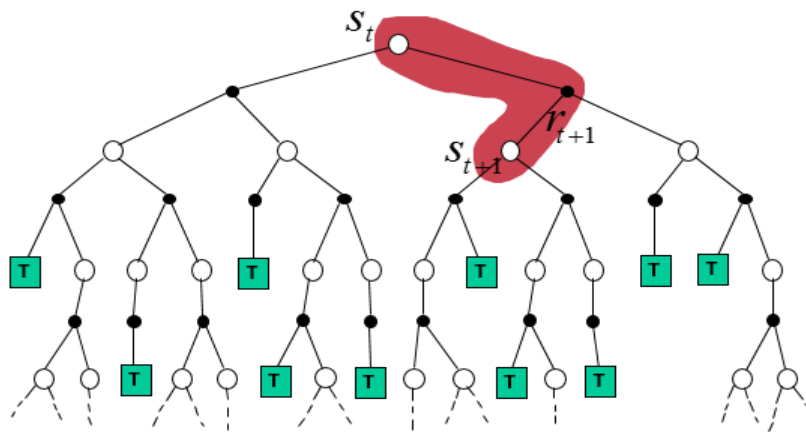
¹ Temporal Difference learning

² Dynamic Programming (DP)

³ Monte Carlo (MC)

$$V(s_t) = V(s_t) + \alpha[R_t + \gamma V(s_{t+1}) - V(s_t)] \quad (3-12)$$

شکل (۳-۶) نیز نشان دهنده شمای کلی به روزرسانی ارزش حالت S می باشد.



شکل ۳-۶: به روزرسانی ارزش در روش TD

دو روش مهم در یادگیری تفاوت گذرا، یادگیری-Q^۱ و یادگیری سارسا^۲ می باشد

۳-۳-۱- الگوریتم یادگیری Q

الگوریتم یادگیری Q توسط واتکینز در سال ۱۹۹۸ برای حل مسایل تصمیم گیری مارکوف با اطلاعات ناقص معرفی شد [۵۵]. این الگوریتم نیاز به مدل دقیق و شفاف محیط ندارد و می تواند از طریق تجربه ای که در اثر تعامل با محیط به دست می آورد به صورت بهنگام برای یافتن استراتژی بهینه استفاده شود از این رو این روش در گروه الگوریتم های مستقل از شناخت مدل قرار می گیرد. شکل (۳-۷) که نشان دهنده توالی حالت-عمل می باشد را در نظر بگیرید:



شکل ۳-۷: توالی حالت-عمل

^۱ Q-learning

^۲ Sarsa

همانطور که از شکل یک نیز مشخص است پاداش عددی r_t به یک زوج (s_t, a_t) تعلق دارد بنابراین مقدار این پاداش تابعی از حالتی است که سیستم در آن قرار دارد و عملی است که عامل در این حالت انجام می‌دهد. در یادگیری Q تابع مقدار با استفاده از یک جدول حالت-عمل مشخص می‌شود که مقدارهای آن به صورت Q تعریف می‌شود. فرض کنید $S = \{s_1, s_2, \dots, s_k\}$ و مجموعه‌ای از k حالت قابل قبول محیط باشد و $A = \{a_1, a_2, \dots, a_m\}$ مجموعه‌ای از m عمل قابل قبول باشد که توسط عامل قابل انجام است.

در اپیزود t ام عامل:

۱. حالتی که در آن قرار دارد را درک می‌کند.

۲. یک عمل $a_t \in A$ را بر اساس خط‌مشی خود انتخاب کرده و اجرا می‌کند.

۳. حالت جدیدی که به آن می‌رسد را مشاهده می‌کند.

۴. پاداش لحظه‌ای r_t ناشی از این حرکت را دریافت می‌کند.

۵. با توجه به رابطه‌ی زیر مقدار Q جدول را به روز می‌کند.

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha [r_t + \gamma \max_{a'} Q_t(s', a) - Q_t(s, a)] \quad (۳-۱۳)$$

در رابطه فوق s' نشان دهنده حالت بعدی می‌باشد. مطابق رابطه (۳-۱۳) تنها مقدار متعلق به زوج (s, a) تغییر می‌کند و رابطه بالا حداکثر پاداش مورد انتظار دریافتی در نتیجه انجام بهترین عمل در حالت بعدی محیط می‌باشد. ضریب α نرخ یادگیری است که میزان یادگیری عامل از اطلاعات جدید را نشان می‌دهد و در بازه $[0, 1]$ قرار دارد و انتخاب مقدار کوچک α به معنی بهره برداری کم از اطلاعات جدید به دست آمده از آخرین عمل انجام شده می‌باشد و مقدار بزرگ α نتیجه عکس دارد. پاداش دریافتی در حالت بعدی با نرخ کاهشی γ در زمان حال محاسبه می‌شود. جهت در نظر گرفتن

ارزش زیاد برای پاداش مورد انتظار در آینده باید مقدار ضریب γ را افزایش و در غیر اینصورت کاهش

$$\text{داد } \gamma \in [0, 1]$$

یکی از مهمترین گزینه‌ها در الگوریتم یادگیری Q مصالحه بین بهره‌وری از اطلاعاتی که تاکنون به دست آمده و اطلاعات و تجربیات آتی می‌باشد. روش‌های تجربی مانند روش ϵ -حریصانه و استراتژی تصادفی می‌تواند در جهت برقراری مصالحه فوق بکار گرفته شوند. در روش ϵ -حریصانه که در این تحقیق به کار گرفته شده است، عامل با احتمال $1 - \epsilon$ عمل متناظر با ماکزیمم مقدار Q برای حالت معلوم و با احتمال ϵ از میان سایر اعمال یکی را به طور تصادفی انتخاب می‌کند. با استفاده از این روش عامل با جستجوی بیشتر محیط، اطمینان خود را در دستیابی به جواب بهینه افزایش می‌دهد.

تعیین ساختاری مناسب در سیستم‌های پیچیده و غیرخطی مانند سیستم‌های بیولوژیکی سخت است و قاعدتا نمی‌توان از راه‌هایی مانند حدس و آزمون خطا به نتیجه مناسب رسید. تفاوت روش‌های یادگیری مختلف، در مدل‌سازی فرآیند یادگیری عامل می‌باشد.

از میان روش‌های هوشمند الگوریتم یادگیری تقویتی بدلیل توانایی این روش در تصمیم‌گیری عامل‌ها در محیط‌های ناشناخته مورد توجه قرار گرفته است.

۳-۳-۲ الگوریتم سارسا

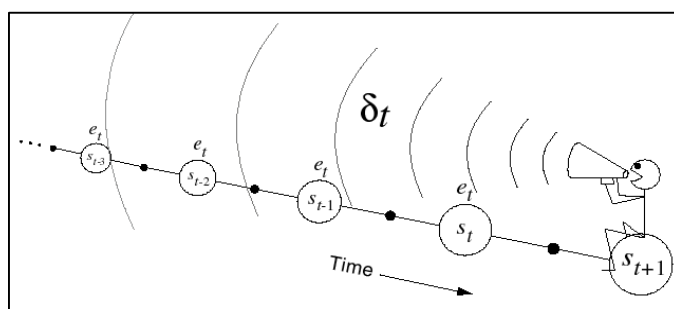
الگوریتم سارسا نیز یکی از روش‌های یادگیری تفاوت گذرا^۱ می‌باشد که مانند الگوریتم یادگیری Q (مانند همه‌ی الگوریتم‌های حل مسئله یادگیری تقویتی) در آن حالت‌ها، عمل‌ها، استراتژی و پاداش تعریف می‌گردد. در الگوریتم سارسا به روز شدن جدول Q به صورت رابطه‌ی (۳-۱۴) می‌باشد [۵۱].

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha [r_t + \gamma Q_t(s', a') - Q_t(s, a)] \quad (3-14)$$

¹ Temporal difference(TD)

۳-۳-۴- مسیرهای شایستگی^۱ (ET)

روش مسیرهای شایستگی تعاملی بین دو روش دیگر حل مسئله‌ی یادگیری تقویتی به نام‌های مونت کارلو^۲ (MC) و یادگیری تفاوت‌گذرا^۳ (TD) می‌باشد. در این روش، به‌روزرسانی ارزش جفت حالت - عمل در هر گام انجام می‌گیرد بدین صورت که برای به‌روزرسانی ارزش حالت و عمل انتخابی فعلی علاوه بر نگاه رو به جلو در هر حالت (s_t)، نگاه به عقب (نگاه دوسویه) نیز وجود خواهد داشت. نگاه دوسویه بدین معنا می‌باشد که برای به‌روزرسانی ارزش جفت حالت و عمل در فعلی، نه تنها ارزش حالت و عمل در لحظه بعد تاثیرگذار است، بلکه ارزش کل زنجیره حالت و عمل‌های قبلی نیز در به‌روزرسانی آن تاثیرگذار خواهد بود. این امر موجب افزایش سرعت در کاهش جمعیت سلول‌های سرطانی با تزریق بهینه‌ترین میزان دوز دارو خواهد شد که کاهش اثرات زیان‌بار دارو را به همراه خواهد داشت [۵۳]. نگاه به عقب در روش مسیرهای شایستگی در شکل (۳-۸) قابل نمایش می‌باشد.



شکل ۳-۸: نگاه به عقب در روش مسیرهای شایستگی [۵۱]

شایستگی زوج حالت و عمل توسط رابطه‌ی (۳-۱۵) به‌روز می‌شود.

$$e_{t+1}(s, a) = e_t(s, a) + 1 \quad (۳-۱۵)$$

به‌روزرسانی سیگنال تقویتی (δ)، طبق رابطه‌ی (۳-۱۶) انجام می‌شود [۵۳].

$$\delta = r + \gamma Q_t(s', a') - Q_t(s, a) \quad (۳-۱۶)$$

در رابطه‌ی (۳-۱۶)، r میزان پاداش دریافتی و $Q(s', a')$ ارزش جفت حالت و عمل در لحظه

^۱ Eligibility Traces

^۲ Monte Carlo

^۳ Temporary Differences Learning

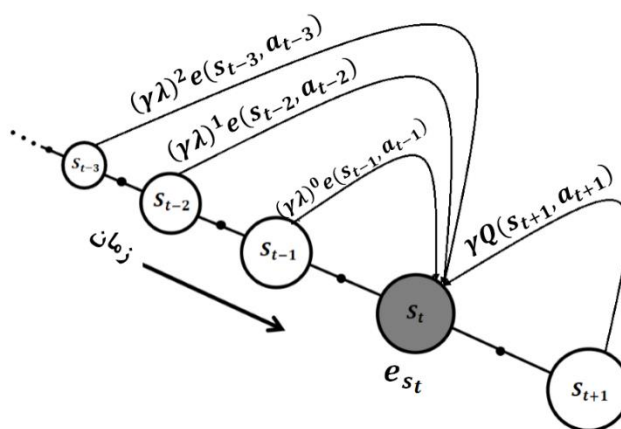
بعدی ($t+1$) می‌باشد. برای بروزرسانی ارزش جفت حالت و عمل از رابطه‌ی (۳-۱۷) استفاده می‌شود [۵۳].

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha \delta e_t(s, a) \quad (۳-۱۷)$$

در رابطه فوق، α نرخ یادگیری می‌باشد. بروزرسانی میزان شایستگی نیز، توسط رابطه (۳-۱۸) محاسبه می‌شود [۵۳].

$$e_{t+1}(s, a) = \gamma \lambda \times e_t(s, a) \quad (۳-۱۸)$$

در رابطه فوق، γ نرخ فراموشی می‌باشد. λ ، پاداش‌ها را از گام فعلی تا انتهای مسیر وزن‌دهی می‌کند و مقدار آن ثابت و بین صفر و یک انتخاب می‌شود. طبق این رابطه، بروزرسانی میزان شایستگی زوج حالت و عمل برای حالت‌های قبلی و حالت فعلی، با بودن در هر حالت (s_t) در مقدار $\gamma \lambda$ ضرب خواهد شد [۵۳]. نمای کلی از ساختار در شکل (۳-۹) نشان داده شده است.



شکل ۳-۹: نمای کلی از روند به روزرسانی مسیرهای شایستگی

فصل ۴- مدل ریاضی بیماری دیابت و سرطان ملانوما (سیستم‌های

مطالعه موردی)

۴-۱- بیماری دیابت

دیابت سندرمی ناشی از هایپرگلیسمی است که باعث عوارض عمده در اکثر اندام‌های بدن می‌شود. چند نوع مشخص و مجزای دیابت شیرین وجود دارند که در اثر واکنش‌های پیچیده‌ای که بین عوامل ژنتیکی و فاکتورهای محیطی رخ می‌دهد، بوجود می‌آیند.

اهمیت روز افزون دیابت چندین دلیل دارد. از جمله اینکه میزان شیوع جهانی دیابت شیرین طی دو دهه گذشته به نحو چشم‌گیری افزایش یافته است. در اینجا این نکته نیز قابل ذکر است که اگر چه میزان شیوع دیابت شیرین نوع ۱ و نوع ۲ در سراسر جهان در حال افزایش می‌باشد ولی سرعت افزایش شیوع دیابت شیرین نوع ۲ بسیار بیشتر است. از طرف دیگر دیابت شایع‌ترین اختلال اندوکراین است و شایع‌ترین علت کوری و مراحل نارسایی کلیوی^۱ (ESRD) و قطع عضو^۲ را هم تشکیل می‌دهد. این بیماری همچنین زمینه‌ساز ابتلا به بیماری‌های قلبی عروقی است. [۵۶].

طبقه‌بندی دیابت شیرین در گذشته بر اساس معیارهای قدیمی‌تری نظیر سن شروع دیابت و یا نوع درمان بود اما امروزه بر اساس روندهای بیماری‌زایی که باعث هایپرگلیسمی می‌شوند، تقسیم می‌شوند. دیابت اولیه شامل ۳ نوع^۳ DM نوع ۱، DM نوع ۲ و دیابت^۴ MODY است. دیابت ثانویه به علت عواملی است که انسولین پانکراس را کاهش می‌دهند مثل پانکراتیت. دیابت بارداری به علت تغییرات متابولیک در اواخر حاملگی است که مقاومت به انسولین افزایش می‌یابد. دیابت شیرین نوع ۱ در نتیجه کمبود کامل یا تقریباً کامل انسولین رخ می‌دهد. دیابت شیرین نوع ۲ شامل گروه

¹ end stage renal disease

² non-traumatic amputation

³ Diabetes mellitus

⁴ maturity onset diabetes of the young

ناهمگونی از اختلالات است که با درجات متفاوتی از مقاومت به انسولین، اختلال ترشح انسولین و افزایش تولید گلوکز مشخص می شوند [۵۶، ۵۷].

برای تشخیص دیابت روش‌های گوناگونی وجود دارد که پایه و اساس این روش‌ها، اندازه‌گیری قند خون است. تحمل گلوکز را به سه گروه اصلی، هموستاز طبیعی گلوکز، اختلال هموستاز گلوکز و دیابت شیرین، می‌توان دسته‌بندی نمود. تحمل گلوکز را می‌توان با استفاده از سطح پلاسمایی گلوکز در حالت ناشتا^(۱) (FBS)، میزان پاسخ به مصرف گلوکز خوراکی^(۲) (OGT) یا میزان هموگلوبین A1C ارزیابی نمود.

به طور کلی درمان دیابت ۳ رکن اصلی دارد: ورزش، رژیم و دارو. دارو درمانی شامل انسولین و داروهای پایین آورنده‌ی قندخون می‌باشد. برای درمان DM تیپ یک فقط انسولین تجویز می‌شود. در دیابت تیپ دو با قند خون زیر ۲۰۰ و بدون عارضه، قبل از شروع درمان دارویی، باید بیمار به مدت ۳ تا ۴ هفته تحت رژیم درمانی و ورزش قرار گیرد و در صورت عدم کنترل پس از ۳-۴ هفته، دارودرمانی شروع می‌شود. در صورتیکه بیمار مورد نظر به درمان دارویی جواب ندهد، انسولین را به رژیم دارویی اش اضافه می‌شود. هر چند که انسولین در چند مورد خاص، درمان اولیه در DM تیپ دو است [۵۶].

افراد مبتلا به دیابت نیاز به کنترل مداوم میزان قند خون دارند. مهم‌ترین هدف در درمان دیابت نگه داشتن گلوکز خون در حد طبیعی می‌باشد. در واقع مطابق شکل ۴-۱ بدست آوردن سیگنال کنترلی بهینه که نرخ تزریق انسولین می‌باشد. مهم‌ترین موضوع در امر تنظیم سطح گلوکز خون در حد طبیعی بدن خواهد بود. مطابق با دیاگرام بلوکی شکل ۴-۱، نرخ تزریق انسولین به عنوان سیگنال کنترلی، توسط پمپ به بیمار دیابتی اعمال می‌شود [۵۸].

¹ Fasting blood sugar

² Oral glucose test



شکل ۴-۱: کنترل حلقه بسته بیماری دیابت با عملگر تزریق پمپ

در شکل ۴-۱ کنترل کننده نرخ میزان تزریق انسولین را مشخص می‌کند. براساس نرخ تزریق تعیین شده، پمپ هیدرولیکی، انسولین را به بدن بیمار تزریق می‌نماید. در نتیجه غلظت گلوکز خون فرد بیمار تغییر می‌یابد. میزان دقیق غلظت گلوکز توسط سنسور محاسبه می‌گردد. سپس اختلاف آن با میزان گلوکز خون طبیعی انسان محاسبه می‌شود و چرخه‌ی فوق تا کنترل کامل سیستم ادامه می‌یابد.

۴-۲- مروری بر اقدامات صورت گرفته جهت کنترل غلظت گلوکز - انسولین بیمار

دیابتی

تاکنون روش‌های کنترلی متعددی جهت کنترل سطح غلظت گلوکز خون بیمار دیابتی ارائه گردیده است. در ادامه به بررسی برخی از آن‌ها خواهیم پرداخت. از طریق ترکیب کننده‌ی PID و روش فازی بهینه یک سیستم خودکار یک پارچه به گونه‌ای طراحی شده است که با استفاده از یک سنسور برای اندازه‌گیری سطح گلوکز خون و یک پمپ کوچک جهت تزریق انسولین می‌تواند نقش سلول‌های بتا در لوزالمعده را ایفا کند [۵۹]. با در نظر گرفتن ماهیت تاخیری مدل دالامن، ساختار

کنترلی جدیدی برای سیستم‌های تاخیردار پیشنهاد گردیده که شامل تلفیقی از کنترل مدل مرجع تطبیقی و پیش‌بین اسمیت اصلاح یافته می‌باشد [۶۰]. در سال‌های اخیر الگوریتم کنترلی متشکل از یک تنظیم‌کننده‌ی ساختار متغیر به همراه پارامترهای کنترل از طریق خطی‌سازی معادلات دیفرانسیل غیرخطی تاخیردار ارائه شده است. این الگوریتم مبتنی بر روش نابرابری ماتریس خطی بوده و با استفاده از قانون کنترلی کلیدزنی، بسته به ارزش فعلی گلوکز و در حضور اختلالات وعده غذایی به کار رفته است [۶۱]. در سال گذشته یک مدل ریاضی جدید برای سیستم گلوکز - انسولین متناسب با داده‌های بالینی دو روزه از افراد سالم ارائه شده است. داده‌های بالینی استاندارد مربوط به مانیتورینگ غلظت گلوکز، تزریق انسولین و مصرف کربوهیدرات از دو دانشگاه نانتس و رنس گرفته شده‌اند [۶۲]. در [۶۳] یک کنترل‌کننده‌ی فازی-ژنتیک جهت کنترل غلظت گلوکز - انسولین ارائه شده است. اخیراً یک کنترل‌کننده‌ی پارامتر متغیر خطی کلیدزنی برای تنظیم گلوکز خون طراحی شده است. هدف از این کنترل‌کننده به حداقل رساندن خطرات ناشی از هیپوگلیسمی و هایپرگلیسمی می‌باشد. این ساختار کنترلی با در نظر گرفتن دینامیک غیرخطی متغیر با زمان و اعمال عدم قطعیت پارامتری به تنظیم سطح گلوکز خون بیمار می‌پردازد [۶۴]. یک ساختار کنترلی برای کاهش خطرات ناشی از افزایش و یا کاهش قند خون در دیابت نوع ۱ به کار رفته است. این ساختار از سه بخش اصلی تشکیل شده است: یک کنترل‌کننده‌ی مقاوم، یک حلقه فیدبک انسولین و یک مکانیزم ایمنی [۶۵]. موارد فوق بخشی از کارهای ارائه شده در زمینه‌ی کنترل بیماری دیابت می‌باشد. در مقاله [۶۶]، از الگوریتم یادگیری Q در مسئله یادگیری تقویتی برای تنظیم میزان سطح گلوکز خون در بیماران مبتلا به دیابت نوع یک استفاده کرده‌اند. عامل یادگیری با جستجو در محیط یاد می‌گیرد که بهترین عمل که میزان دوز انسولین می‌باشد، را انتخاب کند. حالت‌ها، میزان تغییرات هموگلوبین خون در نظر گرفته شده است که در شش حالت طبقه‌بندی شده‌اند. با استفاده از الگوریتم یادگیری Q عامل با دریافت پاداش از محیط، یاد می‌گیرد که بهترین میزان دوز دارو را تزریق کند تا سطح گلوکز خون در بیماران دیابتی روی مقدار مطلوب باقی بماند. در مقاله [۶۷]، از منطق فازی برای کنترل سطح گلوکز و

تزریق انسولین استفاده کرده‌اند. روند کاری در این مقاله به این صورت می‌باشد که در یک سیستم حلقه بسته هر لحظه سطح گلوکز خون اندازه‌گیری شده و بر اساس آن کنترلر منطق فازی بهترین میزان انسولین را پیشنهاد می‌دهد تا میزان گلوکز خون به مقدار مطلوب برسد. در مقاله [۶۸]، با استفاده از کنترلر فازی، تنظیم میزان گلوکز در بیماران دیابتی نوع یک انجام گرفته است. توابع عضویت روی زمان، میزان گلوکز، سرعت و میزان انسولین تعریف شده است. در مقاله [۶۹]، با استفاده از انواع روش‌های داده‌کاوی و یادگیری ماشین برای تنظیم میزان گلوکز در بیماران مبتلا به دیابت استفاده کرده‌اند. از مواردی که در این مقاله استفاده شده و با هم مقایسه شده‌اند، می‌توان به روش ماشین بردار پشتیبان، k -نزدیک‌ترین همسایه و الگوریتم بیز اشاره کرد. آن‌ها نشان داده‌اند که از بین روش‌های پیشنهادی، روش ماشین بردار پشتیبان بهترین دقت را داشته است. در مقاله [۷۰]، با استفاده از روش یادگیری تقویتی، تنظیم میزان گلوکز در بیماران دیابتی نوع دو انجام گرفته است. انجام فعالیت‌های بدنی برای افرادی که مبتلا به دیابت هستند، بسیار مفید می‌باشد. روند کاری این مقاله نیز تشویق افراد مبتلا به دیابت به انجام فعالیت‌های بدنی از جمله راه رفتن می‌باشد. الگوریتم نوشته شده توسط یادگیری تقویتی به این صورت می‌باشد که برای ۲۷ بیمار به صورت هفتگی یک پیامک به تلفن همراه آن‌ها ارسال شده و رژیم غذایی و فعالیت‌های بدنی مورد نیاز به آن‌ها یادآوری می‌شود. نتایج حاصل حاکی از آن می‌باشد که افرادی که تحت این الگوریتم مورد آزمایش قرار گرفته‌اند، کاهش بیشتر میزان قند خون را داشته‌اند. در مقاله [۷۱]، تنظیم میزان انسولین برای کنترل گلوکز در بیماران مبتلا به دیابت نوع یک، با استفاده از الگوریتم یادگیری Q انجام گرفته است. حالت‌ها روی مقدار گلوکز و در پنج بازه تقسیم‌بندی شده است. در نهایت نشان داده شده است که میزان گلوکز به صورت نمایی کاهش پیدا می‌کند. در مقاله [۷۲]، کنترل میزان گلوکز خون در بیماران مبتلا به دیابت با استفاده از یادگیری تقویتی انجام گرفته است. در مقاله مروری [۷۳] روش‌های مدیریت بیماری دیابت با استفاده از روش‌های هوش مصنوعی به صورت جامع مورد بحث قرار گرفته است.

۴-۳- مدل‌های ریاضی بیماری دیابت

برای توصیف دینامیک بیمار مبتلا به دیابت، مدل‌های ریاضی مختلفی وجود دارد [۵۷، ۷۴-۸۱]. دو مدل ریاضی که در این رساله استفاده شده است، در ادامه توضیح داده شده است.

۴-۳-۱- مدل ریاضی دانکومار و سندیا

مدل ریاضی دانکومار و سندیا^۱ ارائه شده است [۸۲]. آن‌ها دینامیک بدن بیمار مبتلا به دیابت را با سه معادله دیفرانسیل توصیف کرده‌اند. این معادلات در روابط (۴-۱) تا (۴-۳) بیان شده‌اند:

$$\frac{dG}{dt} = -m_1 G + m_2 I + m_1 G_b \quad (۴-۱)$$

$$\frac{dX}{dt} = -m_2 X + m_3 I - m_3 I_b + m_6 I_b \quad (۴-۲)$$

$$\frac{dI}{dt} = -m_3 I + m_4 G + m_4 m_5 - m_6 I + m_6 I_b \quad (۴-۳)$$

این مدل مرتبه ۳ شامل متغیرهای G ، گلوکز فرد، X ، متغیر عمومی انسولین برای جز گذشته و I انسولین کنونی فرد می‌باشد. متغیر انسولین قبلی، میزان برداشت انسولین را بین گلوکز خون و بافت‌های محیطی و کبد تنظیم می‌کند. از مزیت‌های این مدل می‌توان به وجود ثوابت G_b و I_b اشاره کرد که نشان‌دهنده مقادیر پایه‌ای برای گلوکز و انسولین می‌باشد. متغیر u در رابطه (۴-۳)، میزان دوز داروی تزریقی به بیمار برای کنترل میزان گلوکز می‌باشد. ثوابت سیستم در جدول (۴-۱) تعریف شده‌اند.

¹ Sandhya danKumar

جدول ۴-۱: ثوابت سیستم دانکومار و سندیا [۸۲]

پارامتر	توضیحات	واحد
G_b	مقدار پایه‌ای گلوکز پلاسمایی قبل از تزریق	$\frac{mg}{dl}$
I_b	مقدار پایه‌ای انسولین پلاسمایی قبل از تزریق	$\frac{\mu U}{ml}$
m_1	میزان بازجذب گلوکز در عضلات و کبد و بافت‌های چربی که مستقل از سطح انسولین می‌باشد.	min^{-1}
m_2	میزان کاهش در نرخ بازجذب گلوکز توسط بافت‌ها	min^{-1}
m_3	میزان افزایش در نرخ بازجذب گلوکز در بافت‌ها (که مستقل از میزان گلوکز می‌باشد) به ازای هر واحد از غلظت انسولین در بالای I_b	$min^{-2}(\frac{\mu U}{ml})$
m_4	میزان رهایش انسولین از سلول‌های B پانکراس، بعد از تزریق گلوکز در حالی که غلظت گلوکز بالای h می‌باشد	$(\frac{\mu U}{ml})min^{-2}(\frac{mg}{dl})$
m_5	میزان سطح آستانه‌ای از گلوکز که بالای آن مقدار سلول‌های B پانکراس انسولین ترشح می‌کنند	---
m_6	نرخ تخریب درجه اول برای انسولین در پلاسما	min^{-1}

۴-۳-۲- مدل ریاضی پالمبو

مدل دیگری که در این رساله استفاده شده است مدل ریاضی ارائه شده توسط پالمبو^۱ می‌باشد [۸۳]. از جمله مزایای این مدل نسبت به سایر مدل‌ها لحاظ نمودن تأثیر تأخیر معادل با فاصله‌ی زمانی بالا رفتن سطح قند خون تا زمانی که ترشح انسولین توسط لوزالمعده صورت می‌گیرد، می‌باشد. هرچند که اعمال تأخیر موجب پیچیدگی رفتار سیستم و سخت شدن تحلیل آن می‌شود ولی باعث افزایش دقت و نزدیک شدن به مکانیزم طبیعی سیستم می‌شود. از برتری این روش نسبت به روش‌های تأخیری دیگر قابلیت سازگاری آن با آزمایشات بالینی صورت گرفته بر روی افراد سالم و تطبیق داشتن آن با مفاهیم فیزیولوژیکی و کاهش پیچیدگی روابط در عین عملکرد مطلوب می‌باشد. معادلات دینامیکی مدل پالمبو تأخیری به صورت زیر می‌باشد.

$$\frac{dG(t)}{dt} = -K_{xgi}I(t)G(t) + \frac{T_{gh}}{V_g} \quad (۴-۴)$$

$$\frac{dI(t)}{dt} = -K_{xi}I(t) + \frac{T_{igmax}}{V_i} f\left(G(t - t_g)\right) + u(t) + d(t) \quad (۴-۵)$$

¹ Palumbo

$$f(G) = \frac{(\frac{G}{G^*})^\gamma}{1 + (\frac{G}{G^*})^\gamma} \quad (۴-۶)$$

در روابط بالا $G(t)$ و $I(t)$ به ترتیب نشان‌دهنده‌ی غلظت گلوکز خون موجود در پلاسما و غلظت انسولین موجود در پلاسما هستند. $f(G)$ یک تابع غیرخطی جهت مدل کردن تاثیر اعمال تاخیر زمانی افزایش سطح گلوکز خون بر نرخ تحویل انسولین از لوزالمعده است و $u(t)$ نیز ورودی کنترلی می‌باشد. در فرمول (۴-۶)

G^* میزانی از گلوکز پلاسما می‌باشد که در آن ترشح انسولین به نیمی از مقدار ماکزیمم خود رسیده است و

γ عکس‌العمل پانکراس در برابر گلوکز موجود در پلاسما را نشان می‌دهد. ثوابت سیستم در جدول (۴-۲) تعریف شده‌اند.

جدول ۴-۲: معرفی پارامترهای مربوط به روابط مدل ریاضی پالمبو تاخیری [۸۳]

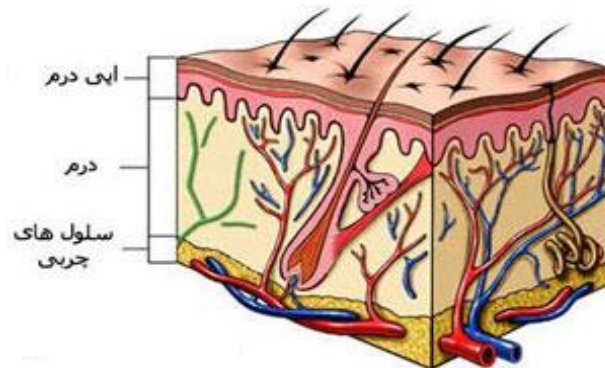
پارامتر	تعریف	واحد
K_{xgi}	نرخ جذب گلوکز خون وابسته به انسولین	$\text{min}^{-1} \text{pM}^{-1}$
T_{gh}	شاخص تعادل میان گلوکز کبدی و گلوکز مستقل از انسولین مرتبه صفر جذب شده توسط بافت‌ها	$\text{min}^{-1} (\frac{\text{mmol}}{\text{kgBW}})$
V_g	نرخ توزیع گلوکز	$\frac{L}{\text{kgBW}}$
K_{xi}	نرخ ناپدید شدن انسولین مرتبه اول	min^{-1}
T_{igmax}	حداکثر میزان تزریق انسولین در فاز دوم	$\text{min}^{-1} (\frac{\text{mmol}}{\text{kgBW}})$
V_i	نرخ توزیع انسولین	$\frac{L}{\text{kgBW}}$
t_g	تأخیر موجود در ترشح انسولین از پانکراس در ازاء افزایش غلظت گلوکز خون که بر حسب دقیقه در نظر گرفته شده	min

۴-۴ - سرطان ملانوما^۱

یکی از وخیم‌ترین انواع سرطان پوست، ملانوما می‌باشد شیوع بیشتر این بیماری در آمریکا بوده به طوری که در هر سال حدود ۵۳۶۰۰ نفر در اثر ابتلا به این بیماری جان خود را از دست می‌دهند. معمولاً سن افرادی که به این بیماری مبتلا می‌شوند بالای ۳۰ سال می‌باشد [84].

ابتدا به تعریف پوست و انواع لایه‌های آن پرداخته می‌شود. پوست یکی از بزرگ‌ترین اعضا بدن است. که یک سد محافظتی بین محیط جهان اطراف با عضلات، اعضای درونی و رگهای خونی و اعصاب بدن تشکیل می‌دهد. پوست از سه بخش اصلی تشکیل شده است:

(۱) اپیدرم (بیرونی ترین لایه پوست می‌باشد). (۲) درم (در زیر اپیدرم قرار گرفته است. و شامل رگهای خونی، لنفاوی و غده‌ها می‌شود). (۳) سلول‌های چربی (این لایه پوست را به بافت‌های ماهیچه‌ای متصل می‌نماید) این سه لایه پوست در شکل (۴-۲) قابل نمایش می‌باشد.



شکل ۴-۲: نمایش سه لایه پوست

۴-۵ - سرطان پوست

در هر روز به طور عادی سلول‌های پوست بدن هر فرد پیر شده و سلول زنده جایگزین آن‌ها می‌گردند. به هم خوردن این روند باعث بروز سرطان پوست خواهد شد. بدین صورت که سلول‌های جدید در هنگام نیاز پوست به آن‌ها به وجود نمی‌آیند و سلول‌های پیر نیز در زمانی که باید نمی‌میرند. این امر

¹ Melanoma Cancer

باعث بوجود آمدن سلول‌های اضافه و ایجاد یک توده‌ای از بافت می‌گردد که تومور نام دارد. این تومورها در دو نوع خوش‌خیم و بدخیم می‌باشند [۸۵].

تومورهای خوش‌خیم دارای سه ویژگی اصلی می‌باشند: می‌توان آن‌ها برداشت و قابلیت رشد مجدد ندارند، عدم حمله به بافت‌های مجاور، عدم انتشار به دیگر قسمت‌های بدن.

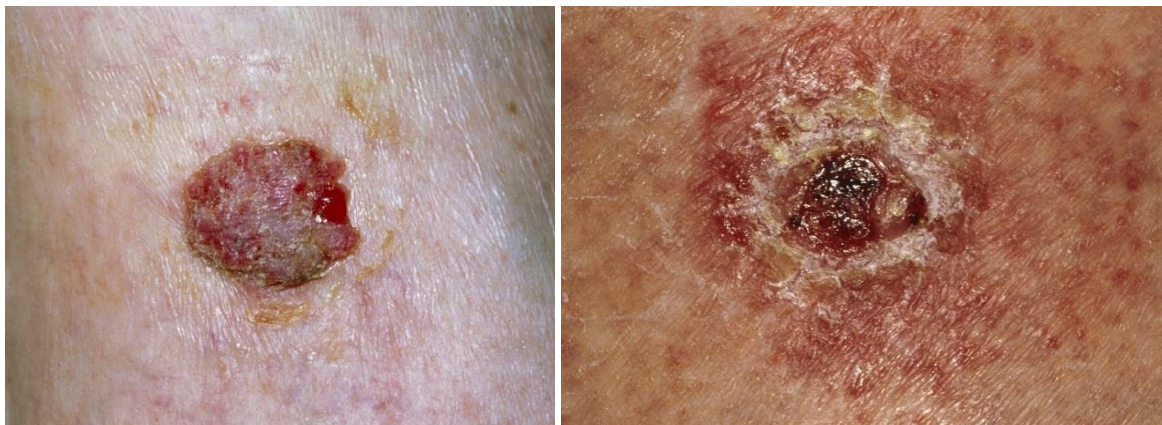
تومورهای بدخیم دارای سه ویژگی اصلی می‌باشند: رشد مجدد حتی بعد از برداشتن تومور، آسیب به بافت‌های مجاور، گسترش به دیگر بخش‌های بدن.

۴-۵-۱- انواع سرطان پوست

سرطان پوست بر اساس نوع ضایعه به سه دسته اصلی تقسیم‌بندی می‌شود:

- سلول‌های سرطانی بنیادی یا پایه (سلول‌های بارازل)

این نوع سرطان در لایه اپیدرم پوست قرار دارد. علایم هشداردهنده آن عبارت است از: زخم‌های مزمن روی پوست، لکه‌های قرمز رنگ همراه با خشکی، درد و خارش که بهبود نمی‌یابند، برآمدگی‌های شفاف‌ی که به رنگ صورتی، قرمز و یا سفید باشند. تصویری از این نوع سرطان در شکل (۳-۴) قابل نمایش می‌باشد [۸۵].



شکل ۴-۴: تصویری از سلول‌های سرطانی فلسی پوست (سلول‌های سنگ فرشی)

شکل ۳-۴: تصویری از سلول‌های سرطانی بنیادی یا پایه (سلول‌های بارازل) [۸۵]

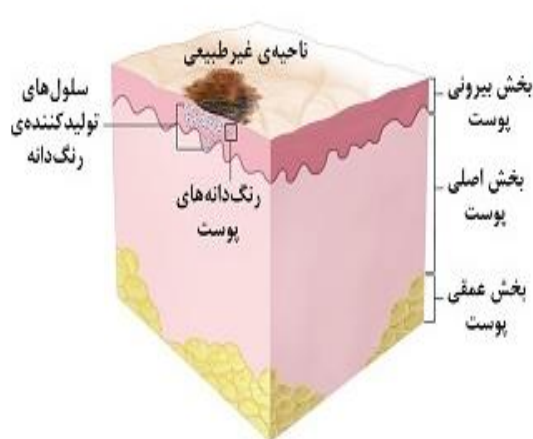
- سلول‌های سرطانی فلسی پوست (سلول‌های سنگ فرشی)

این نوع سرطان تمام بدن را دربر می‌گیرد. علایم هشداردهنده این نوع سرطان عبارت است از:

ضایعه ای شبیه به زگیل زخم باز که خشک شده و خونریزی کند و بهبود نیابد. زخم شبیه به توده‌ای برآمده باشد که خیلی سریع رشد کند. تصویر یاز این سرطان در شکل (۴-۴) نشان داده شده است [۸۵].

- سرطان ملانوما

یکی از بدخیم‌ترین انواع سرطان پوست ملانوما می‌باشد که از رشد بیش از اندازه سلول‌های رنگ‌ساز پوست ایجاد می‌شود و با انتشار به دیگر قسمت‌های بدن می‌تواند باعث مرگ شود. این نوع سرطان معمولاً در رنگ‌های قهوه‌ای روشن و سیاه دیده می‌شود. تصویری از این نوع سرطان پوست و نحوه تشکیل آن در شکل‌های (۴-۵) و (۴-۶) دیده می‌شود [۸۵].



شکل ۴-۶: نمایی از نحوه‌ی تشکیل ضایعه ملانوما.

شکل ۴-۵: تصویری از سرطان ملانوما

این نوع سرطان بیشتر در سنین ۱۰ تا ۱۸ سال اتفاق می‌افتد و دلیل اصلی وقوع آن قرار گیری فرد در معرض تابش مستقیم نور خورشید می‌باشد. نقص در سیستم ایمنی بدن و فاکتورهای ژنتیکی در وقوع این سرطان دخیل می‌باشند. افرادی که بیشتر به این سرطان مبتلا می‌شوند، دارای چنین ویژگی‌هایی می‌باشند: سابقه ابتلا به ملانوم را در قسمتی از بدن خود داشته‌اند، بستگان آن‌ها (پدر،

مادر، خواهر، برادر) به ملانوما مبتلا شده‌اند، رنگ موی آن‌ها قرمز یا بلوند بوده و یا چشم‌های آبی و سبزدارند، سابقه آفتاب سوختگی دارند، افراد بالای چهل سال. لازم به ذکر می‌باشد که خطر ابتلا شدن به ملانوم در مردان بیشتر از زنان می‌باشد.

۴-۶- تشخیص سرطان ملانوما با استفاده از روش‌های پزشکی

یکی از این روش‌ها با استفاده از چک کردن چهار مورد توسط پزشکان انجام می‌گیرد:

۱- غیرقرینگی: قرینه‌گی در ضایعه دیده نمی‌شود.

۲- نامنظمی در حاشیه ضایعه: حاشیه‌های ضایعه به صورت بریده بریده و یا محو باشد.

۳- رنگ: رنگ ضایعه به صورت سایه روشن قهوه‌ای و سیاه می‌باشد و یکنواخت نیست.

۴- اندازه: اندازه ضایعه ملانوم معمولاً بزرگتر از شش میلی‌متر می‌باشد، ولی در اندازه‌های کوچکتر هم احتمال بروز ملانوم وجود دارد.

این روش به دلیل سادگی از اطمینان بالایی برخوردار نیست. به همین خاطر پزشکان از دیگر روش‌ها نیز جهت تشخیص درست استفاده می‌کند. یکی از این روش‌ها خراش روی پوست و گرفتن نمونه خون می‌باشد. ام این روش نیز با درد بسیار برای بیمار همراه است و در بعضی مواقع ممکن است به دلیل عدم تشخیص سریع باعث مرگ بیمار شود.

یکی از روش‌های درمانی برداشتن ضایعه و انجام عمل جراحی می‌باشد. اما احتمال برگشت بیماری نیز وجود دارد. به همین دلیل بیمار بعد از عمل می‌بایست دارو دریافت کند.

۴-۷- مدل ریاضی سرطان ملانوما

مدل‌های ریاضی زیادی برای بیان دینامیک بیماری سرطان ملانوما وجود دارد که در مراجع [۸۶-۹۱] توضیحات کامل مربوط به هر یک از آن‌ها آورده شده است. مدل ریاضی که در این مقاله استفاده شده

است، توسط مارزیو پنیزی^۱ ارایه شده است [۹۲]. یکی از مزایای این مدل ریاضی در نظر گرفتن یک معادله دیفرانسیل جدا برای نمایش رفتار سلول‌های سرطانی می‌باشد. از دیگر مزایای این مدل می‌توان به در نظر گرفتن تعداد متغیرهای بیشتر اشاره کرد. این امر باعث نزدیکی هر چه بیشتر مدل ریاضی به دینامیک بدن بیمار واقعی خواهد شد. این مزیت در دیگر مدل‌های ریاضی از بیمار مبتلا به ملانوما دیده نشده است. متغیرهای در نظر گرفته شده در این مدل از موارد مهمی می‌باشند که در سیستم دفاعی بدن بیمار مبتلا به ملانوما بسیار تاثیرگذار هستند. در این مدل ریاضی برای توصیف دینامیک بدن بیمار از دو محفظه استفاده شده است که نام آن‌ها: محفظه نقطه تزریق^۲ و محفظه پوست^۳ می‌باشد. معادلات بیان‌کننده دینامیک بیماری ملانوما در روابط (۴-۷) تا (۴-۱۳) بیان شده‌اند:

$$\frac{dE}{dt} = K_{in}(t, p) - \alpha_{11}E - \alpha_8E \quad (۴-۷)$$

$$\frac{dAb}{dt} = K_{in}(t, q) - \alpha_{11}Ab - \alpha_{10}Ab \quad (۴-۸)$$

$$\frac{dE_s}{dt} = \alpha_7 \left[\frac{A_s}{A_s + K_1} \right] E_s + \alpha_{11}E(t - \tau) + \alpha_6NA + \alpha_8E_s \quad (۴-۹)$$

$$\frac{dC}{dt} = (\alpha_1 - \alpha_2 \ln(C)).C - \alpha_3 \left[\frac{A_s + K_2}{A_s + K_3} \right] E_s C \quad (۴-۱۰)$$

$$\frac{dA}{dt} = \alpha_4 \left[\alpha_3 \left[\frac{A_s + K_2}{A_s + K_3} \right] E_s C \right] - \alpha_5A - \alpha_6NA \quad (۴-۱۱)$$

$$\frac{dA}{dt} = \alpha_4 \left[\alpha_3 \left[\frac{A_s + K_2}{A_s + K_3} \right] E_s C \right] - \alpha_5A - \alpha_6NA \quad (۴-۱۲)$$

$$\frac{dA_s}{dt} = \alpha_{11}Ab(t - \tau) - \alpha_9A_sE_s - \alpha_{10}A_s \quad (۴-۱۳)$$

معرفی هر یک از متغیرهای حالت در جدول (۳-۴) بیان شده است.

¹ Marzio Pennisi

² Injection point compartment

³ Skin compartment

جدول ۴-۳: معرفی متغیرهای حالت سیستم

نام متغیر حالت	معرفی متغیر	واحد
E	میزان CTL (OT1 Cytotoxic T Lymphocytes) های فعال در محفظه‌ی نقطه‌ی تزریق	$Cell/Kg$
Ab	میزان آنتی بادی CD137 تزریق شده در محفظه‌ی نقطه‌ی تزریق	mg/ml
E_s	میزان OT1 CTL های فعال در محفظه پوست	$Cell/Kg$
C	جمعیت سلول‌های ملانوما	mm^2
A	میزان آنتی‌ژن‌ها	$\mu g \text{ of protein}$
N	میزان CTL های ساده	$cell/days$
A_s	میزان آنتی‌بادی‌ها در محفظه‌ی پوست	mg/ml

بر اساس جدول فوق، CTL ها سلول‌های ایمنی بدن بوده و به عنوان عامل کشنده برای از بین بردن سلول‌های سرطانی می‌باشند. CTL های ساده در بدن هر فردی وجود دارند و به محض ورود سلول‌های سرطانی فعال می‌شوند و خاصیت کشندگی پیدا می‌کنند و تبدیل به CTL های فعال می‌شوند. روابط (۷-۴) تا (۱۳-۴) معادله‌ی (۷-۴) بیانگر OT1 CTL های فعال در محفظه‌ی نقطه‌ی تزریق می‌باشد و معادله‌ی (۸-۴) نیز نمایشگر تغییرات آنتی‌بادی‌ها در این محفظه می‌باشد. تغییرات OT1 CTL های فعال در محفظه پوست با معادله‌ی (۹-۴) نمایش داده شده است. در این رابطه، تغییرات CTL ها در بیمار مبتلا به سرطان ملانوما در هر لحظه، به تغییرات OT1 CTL ها در چند لحظه قبل و در محفظه نقطه تزریق وابسته می‌باشد، که بخش دوم این معادله بیان‌کننده‌ی همین موضوع می‌باشد [92].

رابطه‌ی (۴-۱۰) تغییرات جمعیت سلول‌های ملانوما را نشان می‌دهد. در معادله‌ی (۴-۱۱)، روند تغییرات آنتی‌ژن‌ها نشان داده شده است. معادله‌ی (۴-۱۲) نمایشگر تغییرات CTL‌های ساده بوده و معادله‌ی (۴-۱۳) نیز تغییرات آنتی‌بادی‌ها در محفظه‌ی پوست را نشان می‌دهد. در این معادله نیز میزان آنتی‌بادی‌ها در لحظه‌ی فعلی به تغییرات آن‌ها در چند لحظه‌ی قبل وابسته می‌باشد. در این مدل ریاضی، برای کاهش جمعیت سلول‌های سرطانی دو نوع دارو در نظر گرفته شده است. نوع اول آن که در معادله (۴-۷) نشان داده شده است، تزریق CTL می‌باشد. نوع دوم آن تزریق یک نوع آنتی‌بادی به نام CD-137 می‌باشد که در معادله (۴-۸) نشان داده شده است. دیگر ثوابت سیستم نیز در جدول (۴-۴) بیان شده‌اند.

جدول ۴-۴: تعریف ثوابت سیستم

پارامتر	معرفی پارامتر	پارامتر	معرفی پارامتر
α_1	ها CTL نرخ رشد	α_{11}	سرعت حرکت از محفظه‌ی تزریق به محفظه‌ی پوست
α_2	ها CTL نرخ کاهش	h	ها به وسیله غده CTL نرخ تزریق دوباره از تیموس
α_4	نرخ حذف آنتی‌بادی‌ها در محفظه‌ی پوست به ها CTL وسیله‌ی	K_2	مینیمم آستانه نرخ نابودی سلول‌های سرطانی
α_5	نرخ از هم پاشیده شدن آنتی‌بادی‌ها	K_3	ماکزیمم آستانه نرخ نابودی سلول‌های سرطانی
α_6	های فعال در OT1 CTL ها به CTL نرخ تغییر محفظه‌ی پوست	M	در موقعیت‌های بی‌خطر ها CTL تعداد
α_7	های فعال OT1 CTL حداکثر نرخ تکرار	C_0	مقدار اولیه سلول‌های سرطانی
α_8	های فعال در OT1 CTL نرخ مرگ طبیعی محفظه‌ی پوست و نقطه تزریق	P	مقدار داروی تزریق شده به های فعال در محفظه‌ی تزریق OT1 CTL
α_9	نرخ مرگ آنتی‌بادی‌ها در محفظه‌ی	Q	مقدار داروی تزریق شده به
α_{10}	سرعت از بین رفتن آنتی‌بادی در دو محفظه	τ	تاخیر

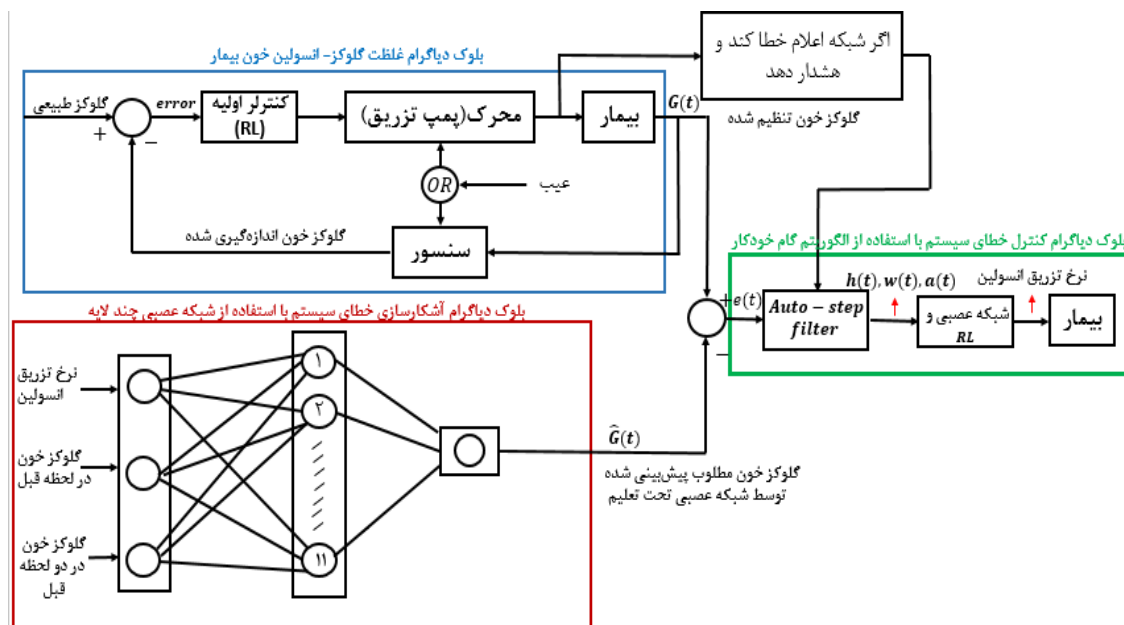
مدل‌های بسیاری در رابطه با بیماری دیابت و سرطان ملانوما مطرح شده است که همانطور که بیان شد دلیل انتخاب این مدل‌ها نزدیکی بیشتر رفتار این مدل‌های ریاضی به بیمار واقعی با توجه به نظرات پزشک متخصص می‌باشد.

فصل ۵- تشخیص عیب به کمک تولید مانده و کنترل تحمل‌پذیر

عیب با ترکیب یادگیری تقویتی و الگوریتم گام خودکار

۵-۱- مقدمه

در این فصل اولین نوآوری و پیشنهاد رساله مطرح گردیده است. در ابتدا میزان گلوکز یک بیمار دیابتی با استفاده از الگوریتم‌های حل مساله یادگیری تقویتی (الگوریتم یادگیری-Q) کنترل می‌شود. محرک و سنسور در این بخش اعداد ثابت فرض شده‌اند. مدل سیستم حلقه بسته نامی در عدم حضور عیب با استفاده از یک شبکه عصبی پیش‌رو زمان گسسته شناسایی گردیده است. در گام بعدی به سنسور و محرک سیستم عیب جمع شونده ناگهانی وارد شده است با مقایسه مقدار واقعی گلوکز و مقدار تخمین زده شده با استفاده از شبکه عصبی مانده تعیین گردیده است، که این مورد در بخش (۵-۲) با توضیحات بیشتر مطرح خواهد گردید. اگر مانده از آستانه مشخصی بیشتر گردد فرض بر این خواهد بود که محرک و یا سنسور سیستم دچار عیب شده است. مقدار آستانه برای تشخیص رخداد عیب با استفاده از سعی و خطا تعیین گردیده است. بعد از تشخیص عیب، کنترلر دیگری جایگزین کنترلر اولیه خواهد شد که این کنترلر بر اساس یادگیری تقویتی پیوسته (با ترکیب شبکه عصبی) پیاده‌سازی گردیده است. در این بخش از الگوریتم گام خودکار برای تعیین پارامترهای یادگیری تقویتی که در واقع همان نرخ آموزش وزن‌های شبکه عصبی می‌باشد استفاده گردیده است. توضیحات لازم در مورد کنترل تحمل‌پذیر عیب در بخش (۵-۳) مطرح خواهد گردید. شکل (۵-۱) شمای کلی ساختار پیشنهادی را نشان می‌دهد.



شکل ۵-۱: شناسایی و کنترل خطای سیستم کنترل کننده‌ی بیماری دیابت توسط ترکیب شبکه‌ی عصبی و یادگیری تقویتی

۵-۲- تشخیص خطا به کمک تولید مانده با استفاده از شبکه‌های عصبی مصنوعی

در ابتدا برای شناسایی رفتار سیستم حلقه بسته نامی از یک شبکه عصبی پیش‌رو چند لایه بهره گرفته‌ایم برای آموزش شبکه از روش پس انتشار خطا^۱ استفاده شده است. داده‌های آموزشی داده‌های سیستم نرمال (در عدم حضور خطا) می‌باشد. پس از آموزش سپس ورودی و خروجی‌ها را به شبکه اعمال کرده‌ایم و تحت الگوریتم پس انتشار خطا به آن آموزش داده‌ایم. شبکه مدل سیستم را به خوبی در مدت زمان کمی فرا می‌گیرد و به درستی مدل سیستم تخمین زده خواهد شد. این شبکه‌ی تعلیم دیده شده به عنوان مبنا و تشخیص دهنده‌ی عیب سیستم مورد

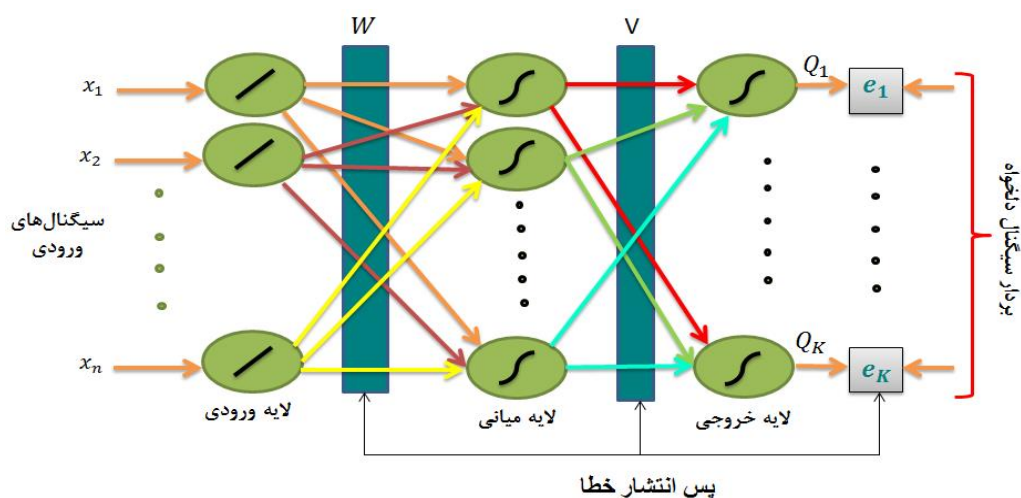
استفاده قرار می‌گیرد و اگر بنا به هر دلیلی در سیستم عیبی رخ دهد که منجر شود انحراف خروجی سیستم از خروجی مورد نظر شبکه بیش از حد معقول باشد؛ آن‌گاه شبکه‌ی عصبی هشدار عیب خواهد داد. تمامی توضیحات فوق را می‌توان در شکل (۵-۱) مشاهده نمود. شبکه عصبی انتخاب شده یک شبکه استاتیکی می‌باشد و برای آنکه بتواند مدل سیستم دینامیکی را شناسایی کند ورودی‌های

¹ Back Propagation

شبکه به صورت سری زمانی از میزان گلوکز و میزان تزریق انسولین در نظر گرفته شده است. برای این امر، میزان گلوکز در پنج لحظه قبل و میزان تزریق انسولین در لحظه فعلی و سه لحظه قبلی لحاظ گردیده است. در بخش بعد مختصری در مورد شبکه عصبی چند لایه و قانون پس انتشار خطا صحبت خواهد شد.

۵-۲-۱- شبکه عصبی پیشرو و الگوریتم پس انتشار خطا

در هر سیستم و تحقیقی، محقق بر حسب نیاز خود می تواند تعداد لایه ها، نرون ها، ورودی و خروجی ها و سایر پارامترها را تغییر و دوباره طراحی نماید. طبق شکل فوق در این مساله شبکه ی مورد نظر دارای سه لایه، دو لایه ی میانی و یک لایه ی خروجی می باشد (معمولا در بسیاری از متون، لایه داده ها را به عنوان لایه در نظر نمی گیرند). ورودی های لایه ی اول همان ورودی های اصلی سیستم جهت تعلیم شبکه هستند که در شکل مشاهده می کنید. خروجی های لایه ی اول، ورودی های لایه دوم یا همان لایه ی میانی می باشند و به همین ترتیب خروجی های لایه ی میانی، ورودی های لایه ی آخر و خروجی های نهایی همان خروجی سیستم خواهد بود که توسط شبکه ی عصبی تخمین زده شده است. این خروجی در شکل (۵-۱) با خروجی سیستم واقعی جهت تشخیص عیب مقایسه می گردد. شکل (۵-۲) ساختار کلی شبکه عصبی پیشرو را نشان می دهد.



شکل ۵-۲: نمایش شبکه عصبی پیشرو

برای به روز رسانی وزن‌های شبکه عصبی به کمک الگوریتم پس انتشار خطا سه گام خواهیم داشت که روابط به صورت زیر خواهد بود [۹۳]:

۱- انتشار به سمت جلو:

$$a^{m+1} = f^{m+1}(w^{m+1}a^m + b^{m+1}) \quad (5-1)$$

در رابطه فوق، $m=0,1,\dots,M$ شماره لایه را مشخص می‌کند. M معرف تعداد لایه‌ها می‌باشد. a^m خروجی لایه m ام و a^{m+1} خروجی لایه $m+1$ ام می‌باشد.

۲- پس انتشار^۱:

حساسیت برای لایه آخر را با استفاده از رابطه (۲-۵) محاسبه می‌شود.

$$S^M = -2\dot{F}^M(n^M)(t - a) \quad (5-2)$$

حساسیت برای دیگر لایه‌ها توسط رابطه (۳-۴) محاسبه می‌گردد.

$$S^m = \dot{F}^M(n^M)(w^{m+1})s^{m+1}; \quad m=M-1, M-2, \dots, 1 \quad (5-3)$$

در هر دو رابطه فوق، t معرف مقدار تارگت^۲ می‌باشد. $\dot{F}^M$ یک ماتریس قطری می‌باشد که با استفاده از رابطه (۴-۵) تعریف می‌شود.

$$\dot{F}^M = \begin{bmatrix} \dot{f}^m(n_1^m) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \dot{f}^m(n_{sm}^m) \end{bmatrix} \quad (4-5)$$

در رابطه فوق، f ماتریسی مشتق‌پذیر بوده و طبق رابطه (۵-۴) محاسبه می‌گردد.

$$\dot{f}^m(n_j^m) = \frac{\partial f(n_j^m)}{\partial n_j^m} \quad (5-5)$$

¹back ward propagation

²target

۳- بروز رسانی وزن ها و بایاس ها:

$$w^m(k+1) = w^m(k) - \alpha s^m(a^{m-1})^T \quad (5-6)$$

$$b^m(k+1) = b^m(k) - \alpha s^m \quad (5-7)$$

۵-۳- کنترل تحمل پذیر عیب به کمک الگوریتم گام خودکار و یادگیری تقویتی

پس از آنکه مدل سیستم حلقه بسته نامی به کمک شبکه عصبی شناسایی گردید آنگاه مانده محاسبه می گردد و اگر مانده از میزانی مشخصی بیشتر شود آنگاه به کمک الگوریتم گام خودکار و یادگیری تقویتی کنترل تحمل پذیر عیب انجام خواهد شد. الگوریتم گام خودکار یک الگوریتم تقویتی خودکار پیشرفته بر اساس فیلتر انطباق است. که توسط ساتن و همکاران معرفی شده است [۴۷]. الگوریتم فوق الگوریتمی گام به گام براساس رویکرد نسبی گرادیان است. به نوعی از زیر مجموعه ای از یادگیری تقویتی جهت تنظیم وزن های شبکه ی عصبی استفاده می کند و برای برنامه های کاربردی مانند یادگیری آنلاین، پردازش سیگنال، حذف صدا و غیره مورد استفاده قرار می گیرد. در این تحقیق از آن برای شناسایی و کنترل عیب سیستم کنترلی غلظت گلوکز- انسولین خون بیمار دیابتی استفاده شده است.

این الگوریتم یک بردار از پارامترهای اندازه ی گام را جهت بهبود عملکرد فیلتر تنظیم می کند. بنابراین، این الگوریتم را می توان به عنوان نسل جدیدی از فیلتر^۱ (IDBD) افزایشی [۹۴] با دو اصلاح اصلی در نظر گرفت. IDBD یک الگوریتم یادگیری قدیمی است که پارامترهای یادگیری را به منظور حذف رویکرد متعصبانه از برآوردها و جلوگیری از قرار گرفتن در مینم های محلی، تنظیم می کند. الگوریتم گام خودکار باعث بهبود عملکرد فیلتر، مستقل از پارامتر اندازه می شود که به صورت خلاصه دارای دو اصلاح زیر می باشد.

¹ Incremental-Delta-Bar-Delta

اولین اصلاح را می‌توان ارائه بهترین مقدار برای پارامتر α جهت به‌روزرسانی وزن‌های شبکه در مواقع ضروری مانند رخداد خطا برای سیستم دانست. در صورت رخداد خطا در سیستم، بعضی از پارامترهای سیستم تاثیر بیشتری در بخش خطا دارند که الگوریتم گام خودکار باعث کاهش حساسیت عملکرد فیلتر به این بخش از پارامترهای سیستم می‌شود و این اصلاح بعدی است که الگوریتم گام خودکار باعث می‌شود. با اینکه احتمال نیاز به محاسبه و حافظه‌ی بالا در الگوریتم گام خودکار با توجه به رویکردهای قبلی افزایش می‌یابد، اما قابلیت انعطاف‌پذیری و تطبیق‌پذیری بالای آن و عدم وابستگی به مدل محیط از ویژگی‌های برجسته‌ی آن می‌باشد.

در فیلتر انطباق، سیگنال ورودی به عنوان یک بردار n تایی بصورت $x_{i,k}$ $[x_{1,k}, x_{2,k}, \dots, x_{n,k}]$ اعمال می‌شود. k عنصر زمان می‌باشد و y_k یک خروجی تک و برآورد شده با استفاده از یک توزیع وزن ورودی برای لحظه‌ی k ام است. خطای تخمین به صورت زیر محاسبه می‌شود:

$$e_k = y_k - \sum_{i=1}^n x_{i,k} w_{i,k} \quad (5-8)$$

در رابطه‌ی فوق w پارامتر مربوط به وزن‌های شبکه می‌باشد. از طریق روابط زیر پارامترهای مربوطه به روز رسانی می‌گردد تا شبکه به طور کامل سیستم را کنترل نماید.

$$\alpha_{i,k+1} = \alpha_{i,k} \exp(\mu x_{i,k} h_{i,k}) \quad (5-9)$$

$$w_{i,k+1} = w_{i,k} + \alpha_{i,k+1} e_k x_{i,k} \quad (5-10)$$

$$h_{i,k+1} = (1 - \alpha_{i,k+1}^2) h_{i,k} + \alpha_{i,k+1} e_k x_{i,k} \quad (5-11)$$

در روابط بالا h متغیر کمکی برای حافظه و α پارامتر اندازه‌ی گام می‌باشد. به کمک روابط فوق و مراجعه به [۹۵] می‌توان اطلاعات جامع‌تری از الگوریتم گام خودکار به دست آورد.

همان‌طور که بیان شد در این رساله از الگوریتم فوق برای تشخیص و شناسایی خطای سیستم کنترل کننده‌ی دیابت استفاده شده است. که نتایج حاصل از آن بر روی مدل پالمبو بیان می‌شود. اما یکی از بزرگ‌ترین دغدغه‌ها پس از تشخیص خطا، کنترل سیستم در حضور خطا تا زمان برطرف شدن عیب می‌باشد. با توجه به ویژگی‌های بیان شده برای الگوریتم گام خودکار، یکی از بهترین گزینه‌ها جهت استفاده برای کنترل سیستم در حضور رخداد خطا می‌تواند باشد. به همین دلیل مطابق شکل ۵-۱ پس از آشکارسازی خطا توسط شبکه‌ی عصبی چندلایه، با استفاده از الگوریتم گام خودکار به کنترل سیستم در حضور خطا پرداخته شده است. در شکل ۵-۱ قسمت آشکارسازی خطا توسط شبکه‌ی عصبی به طور کامل بیان شد. اما در قسمت بلوک دیاگرام کنترل خطای سیستم با الگوریتم گام خودکار، مشاهده می‌کنید الگوریتم، ورودی سیستم را که دچار خطا شده است به همراه مقدار خطا دریافت می‌کند و تلاش می‌کند بر اساس آن‌ها برای کنترل بهینه‌ی سیستم پارامترهای مربوطه را با توجه به روابط بیان شده اصلاح نماید و در اختیار شبکه‌ی عصبی جهت کنترل بهینه‌ی سیستم بگذارد. پس از اقدامات صورت گرفته توسط الگوریتم گام خودکار، سیستم به یک حالت تعادلی پایدار خواهد رسید و به کار خود در حالت ایده‌آل ادامه خواهد داد. برای استفاده از الگوریتم مذکور می‌توان از شبه کد نمایش داده شده در شکل (۵-۳) کمک گرفت.

انجام مقدار دهی اولیه به صورت زیر

$$\eta = 10^4 \text{ و } \mu = 10^{-2}$$

و برای i از ۱ تا n انتخاب مقادیر اولیه ی خیلی کوچک و یا صفر برای v_i و w_i و h_i و α_i صورت گیرد.

در یک حلقه (برای i از 1 تا n)

برای بردار ورودی ها $(x_{i,k} = [x_{1,k}, x_{2,k}, \dots, x_{n,k}])$ ، خطا به صورت زیر محاسبه شود.

$$e_k = y_k - \sum_{i=1}^n x_{i,k} w_{i,k}$$

در یک حلقه (برای i از 1 تا n)

$$V_i = \max(|e x_i h_i|, V_i + \frac{1}{\eta} \alpha_i x_i^2 (I e x_i h_i I - V_i))$$

اگر V_i مخالف صفر شد آنگاه $\alpha_i = \alpha_i \exp(\frac{e x_i h_i}{V_i})$ و $M = \max(\sum_{i=1}^n x_i^2 \alpha_i, 1)$ پایان شرط

$$\alpha_i = \alpha_i / M$$

$$w_i = w_i + \alpha_i e x_i$$

$$h_i = (1 - \alpha_{i,k+1}^2) h_i + \alpha_i e x_i$$

پایان

شکل ۵-۳: یک گام از الگوریتم گام خودکار

شکل ۵-۳ مطالب بیان شده در این فصل را به طور خلاصه و مفید در قالب یک شبه کد بیان می کند. واضح است در هر گام پارامترهای مربوطه اصلاح می گردد تا به بهترین مقادیر، جهت رسیدن به هدف مورد نظر دست پیدا کند.

۵-۴- شبیه سازی

برای نشان دادن کارایی روش پیشنهادی فوق از مدل پالمبو بیماری دیابت استفاده گردیده است. معادلات دینامیکی مدل پالمبو تاخیری به صورت زیر می باشد.

$$\frac{dG(t)}{dt} = -K_{xgi} I(t) G(t) + \frac{T_{gh}}{V_g} \quad (5-12)$$

$$\frac{dI(t)}{dt} = -K_{xi}I(t) + \frac{T_{igmax}}{V_i} f(G(t - t_g)) + u(t) + d(t) \quad (5-13)$$

$$f(G) = \frac{(\frac{G}{G^*})^\gamma}{1 + (\frac{G}{G^*})^\gamma} \quad (5-14)$$

در روابط بالا $G(t)$ و $I(t)$ به ترتیب نشان‌دهنده‌ی غلظت گلوکز خون موجود در پلاسما و غلظت انسولین موجود در پلاسما هستند. $f(G)$ یک تابع غیرخطی جهت مدل کردن تاثیر اعمال تاخیر زمانی افزایش سطح گلوکز خون بر نرخ تحویل انسولین از لوزالمعده است و U نیز ورودی کنترلی می‌باشد. سایر پارامترها به عنوان ثوابت مساله در جدول ۵-۱ نمایش داده شده‌اند.

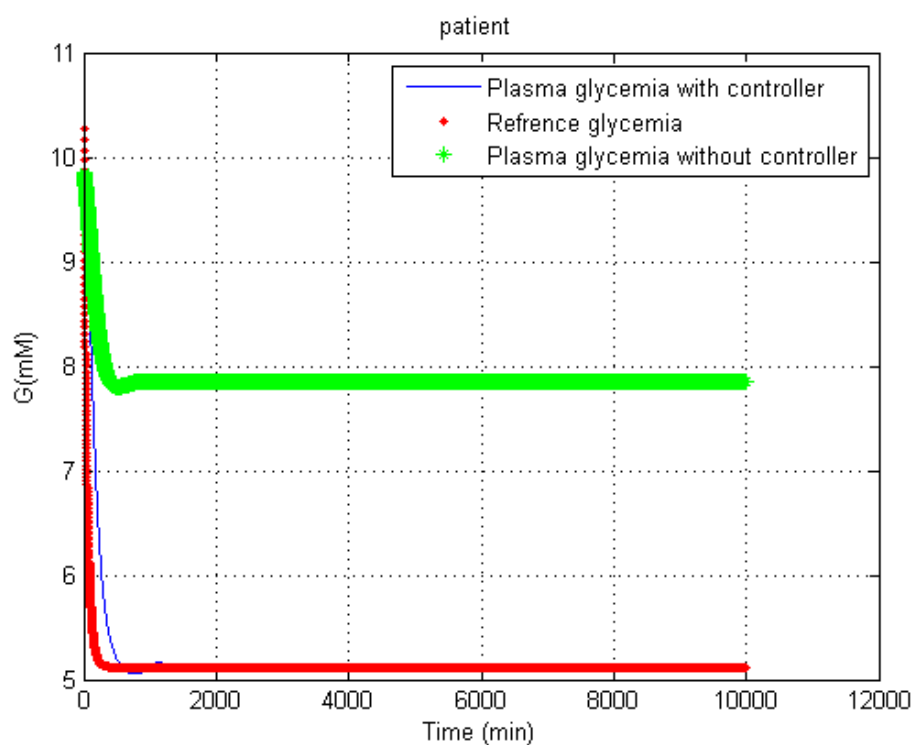
جدول ۵-۱: معرفی پارامترهای مربوط به روابط مدل ریاضی پالمبو تاخیری

پارامتر	تعریف	واحد	مقدار
K_{xgi}	نرخ جذب گلوکز خون وابسته به انسولین	$\text{min}^{-1}\text{pM}^{-1}$	$3.11 \cdot 10^{-5}$
T_{gh}	شاخص تعادل میان گلوکز کبدی و گلوکز مستقل از انسولین مرتبه صفر جذب شده توسط بافت‌ها	$(\text{min}^{-1}(\text{mmol}/\text{kgBW}))$	0.003
V_g	نرخ توزیع گلوکز	L/kgBW	0.187
K_{xi}	نرخ ناپدید شدن انسولین مرتبه اول	min^{-1}	$1.211 \cdot 10^{-2}$
T_{igmax}	حداکثر میزان تزریق انسولین در فاز دوم	$(\text{min}^{-1}(\text{mmol}/\text{kgBW}))$	1.573
V_i	نرخ توزیع انسولین	L/kgBW	0.25
t_g	تأخیر موجود در ترشح انسولین از پانکراس در اِزاء افزایش غلظت گلوکز خون که بر حسب دقیقه در نظر گرفته شده	min	24
G^*	میزانی از گلوکز پلاسما که در آن ترشح انسولین به نیمی از مقدار ماکزیمم خود رسیده است	mM	9
γ	عکس العمل پانکراس در عملکرد گلوکز موجود در پلاسما	-	3.205

۵-۴-۱- شبیه‌سازی در عدم حضور عیب

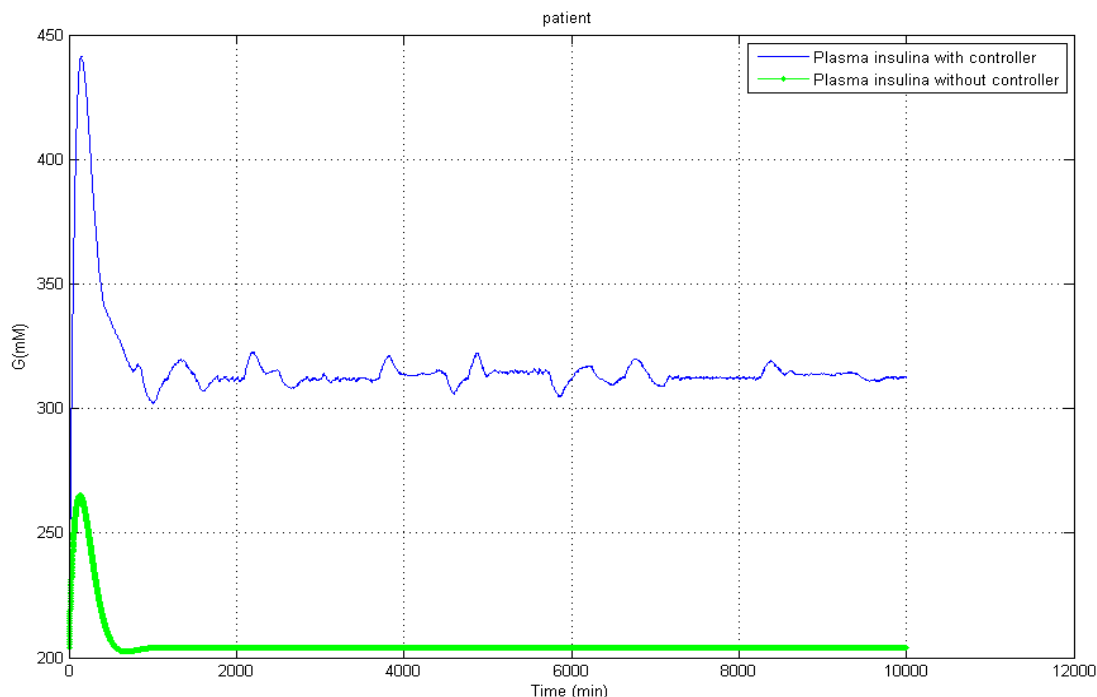
در ابتدا شبیه‌سازی در عدم حضور عیب صورت گرفته است و کنترل‌کننده یادگیری تقویتی توانسته است مقادیر را در حد مطلوب حفظ کند. عامل در این مساله پمپ تزریق انسولین می‌باشد. حالات موجود سیستم ۵ بازه بین $4 \frac{M}{ml}$ تا $6 \frac{M}{ml}$ می‌باشد به صورتی که اولین بازه به اعداد کم‌تر از ۴ و آخرین بازه به اعداد بزرگ‌تر از ۶ تعلق دارد. هر بازه‌ی ذکر شده، نشان دهنده‌ی میزان غلظت گلوکز می‌باشد که طبیعتاً بهترین حالت، نزدیک‌ترین غلظت گلوکز به گلوکز مطلوب که $5 \frac{M}{ml}$ در نظر گرفته شده است می‌باشد. اعمال مساله نیز با توجه به نرمالیزه بودن دوز دارو، ۱۰ عدد بین ۰ و ۱ به عنوان نرخ تزریق انسولین می‌باشد.

عمل تحت سیاست ϵ - حریصانه انتخاب شده است و از الگوریتم یادگیری Q نیز برای آموزش استفاده شده است. برای انجام هر عمل یک پاداش خوب در صورت بهبود عملکرد فرآیند و یک تنبیه در صورت انجام عمل اشتباه در نظر گرفته شده است. پس از پیاده‌سازی مدل و طراحی کنترلر با استفاده از الگوریتم یادگیری تقویتی برای یک بیمار با $G_0 = 10.37 \frac{M}{ml}$ به عنوان غلظت گلوکز خون اولیه و $I_0 = 99.57 \frac{M}{ml}$ به عنوان غلظت انسولین پلاسمای خون اولیه و سایر مشخصاتی که در جدول ۵-۱ می‌باشد؛ نتایج مطلوبی بدست آمد.



شکل ۴-۵: غلظت گلوکز بیمار پس از کنترل توسط کنترل کننده‌ی یادگیری تقویتی

شکل ۴-۵ عملکرد و کاربرد صحیح یادگیری تقویتی در طراحی سیستم کنترل کننده‌ی غلظت انسولین - گلوکز را تایید می‌کند. غلظت گلوکز خون بیمار در حد طبیعی کنترل شده است و سیستم کنترلی به خوبی وظیفه‌ی خود را انجام داده است. مشاهده می‌کنید به مرور زمان، عامل توانسته به خوبی سیستم را کنترل نماید و به غلظت گلوکز خون مرجع برساند. نمودار بالای شکل گلوکز خون بیمار در هنگام عدم حضور کنترل کننده را نمایش می‌دهد.



شکل ۵-۵: غلظت انسولین پلاسمای خون بیمار پس از کنترل

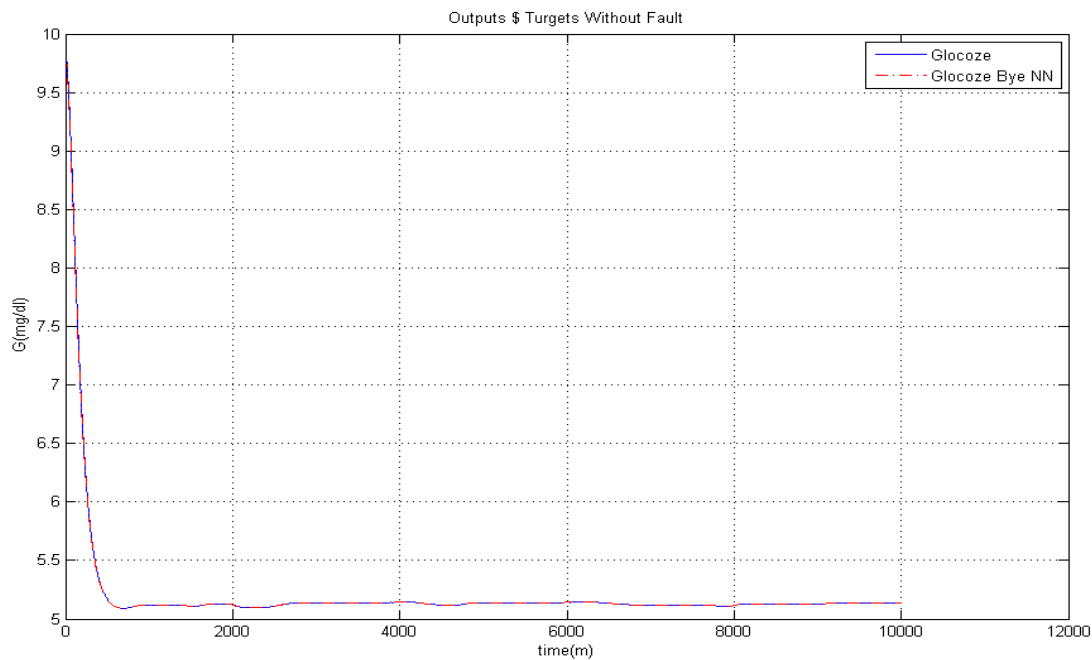
در شکل ۵-۵ مجدداً عملکرد صحیح کنترل کننده تایید می‌شود. ملاحظه می‌گردد غلظت انسولین پلاسمای خون بیمار نیز پس از مدت کوتاهی به حالت ایده‌آل رسیده است و مطابق نمودار بالای شکل در ادامه در همان شرایط خوب کنترل شده است. نمودار پایین شکل، غلظت انسولین پلاسمای خون بیمار در عدم حضور کنترل کننده را نشان می‌دهد.

۵-۴-۲- شبیه‌سازی رفتار کنترل کننده پیشنهادی در حضور عیب

ابتدا ساختار شبکه‌ی عصبی چندلایه طراحی گردیده است. سپس به کمک الگوریتم پس انتشار خطا و رابطه بیان شده برای آن شبکه آموزش دیده است. شبکه عصبی در نظر گرفته شده یک شبکه عصبی دو لایه می‌باشد که لایه اول آن شامل ۸ نرون می‌باشد و لایه خروجی با توجه به آنکه فقط یک خروجی داریم شامل یک نرون می‌باشد توابع فعال‌ساز لایه اول تانژانت سیگموئید^۱ و تابع فعال‌ساز لایه دوم خطی در نظر گرفته شده است. تعداد نرونها لایه پنهان با سعی و خطا و با توجه به تعداد ویژگی‌های ورودی شبکه تعیین گردیده است. همانطور که بیان شد شبکه عصبی انتخاب شده یک

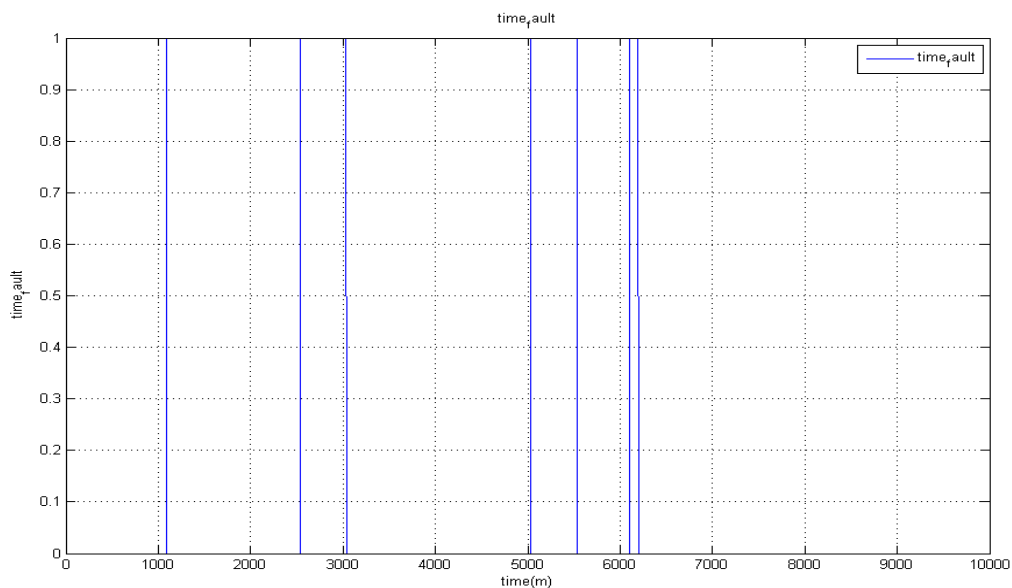
¹ tansig

شبکه استاتیکی می‌باشد و برای آنکه بتواند مدل سیستم دینامیکی را شناسایی کند ورودی‌های شبکه به صورت سری زمانی از میزان گلوکز و میزان تزریق انسولین در نظر گرفته شده است. برای این امر، میزان گلوکز در پنج لحظه قبل و میزان تزریق انسولین در لحظه فعلی و سه لحظه قبلی لحاظ گردیده است.



شکل ۵-۶: تخمین سیستم توسط شبکه‌ی عصبی چندلایه

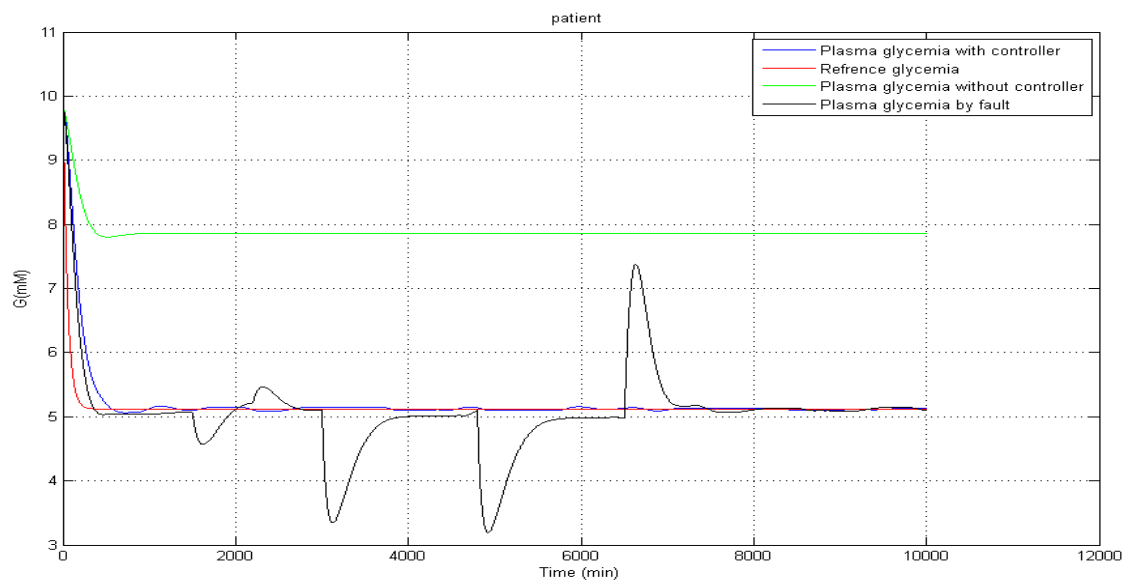
طبق شکل ۵-۶، شبکه به خوبی توانسته است مدل سیستم را فراگیرد و با دقت بالا سیستم را شناسایی کند. اکنون شبکه‌ی عصبی همانند مدل سیستم حلقه بسته نامی رفتار می‌کند. اگر هر ورودی به عنوان میزان نرخ تزریق انسولین به آن تحویل بدهیم؛ خروجی یا همان میزان غلظت گلوکز خون فرد پس از تزریق را به ما می‌دهد.



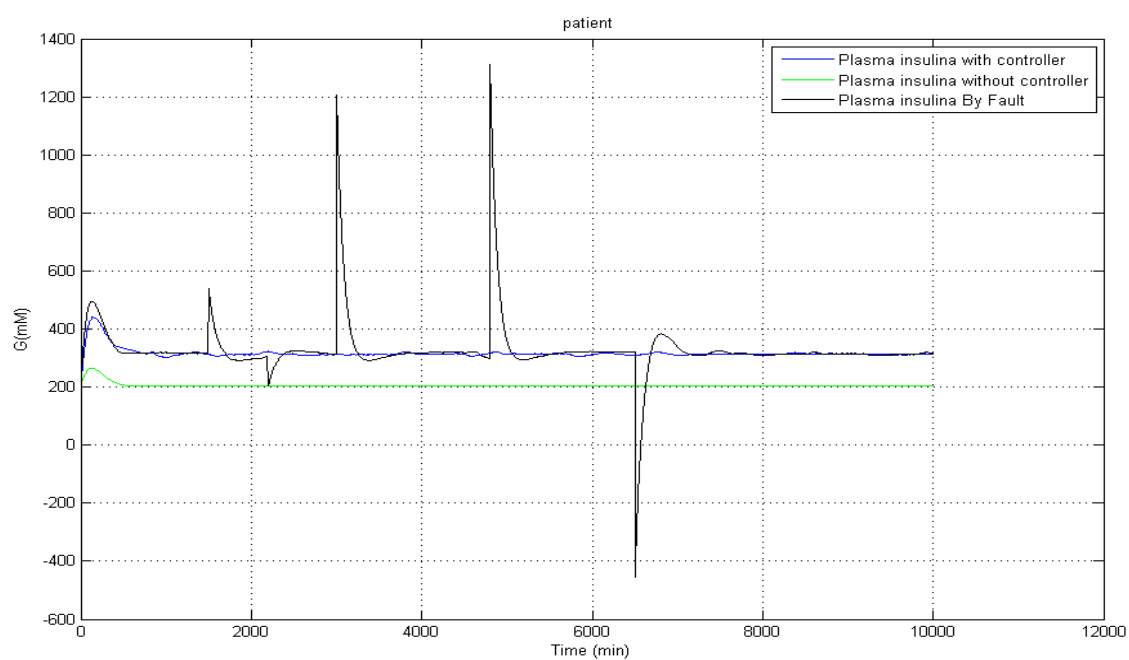
شکل ۵-۷: آشکارسازی خطای سیستم با روش شبکه‌ی عصبی

برای آنکه رفتار سیستم تشخیص عیب مورد ارزیابی قرار گیرد در بازه‌های زمانی مشخص عیب ناگهانی به سیستم اعمال شده است و همانطور که در شکل ۵-۷ ملاحظه می‌کنید سیستم تشخیص عیب طراحی شده مطابق انتظار، زمان‌هایی که سیستم و فرایند، دچار عیب شده است را تشخیص داده و اعلام هشدار نموده است. در تصویر فوق بازه‌های زمانی که سیستم دچار عیب بوده است نمودار روی عدد یک و در سایر زمان‌ها روی عدد صفر قرار داشته است.

شکل (۵-۸) و شکل (۵-۹) نیز به ترتیب رفتار سیستم را در حضور و عدم حضور کنترل کننده در هنگام رخداد عیب نشان می‌دهند.



شکل ۵-۸: میزان تغییرات گلوکز خون در حضور عیب در حضور کنترلر پیشنهادی و در عدم حضور آن



شکل ۵-۹: میزان تغییرات انسولین خون در حضور عیب در حضور کنترلر پیشنهادی و در عدم حضور آن

فصل ۶- کنترل تحمل‌پذیر عیب با استفاده از ترکیب مسیرهای

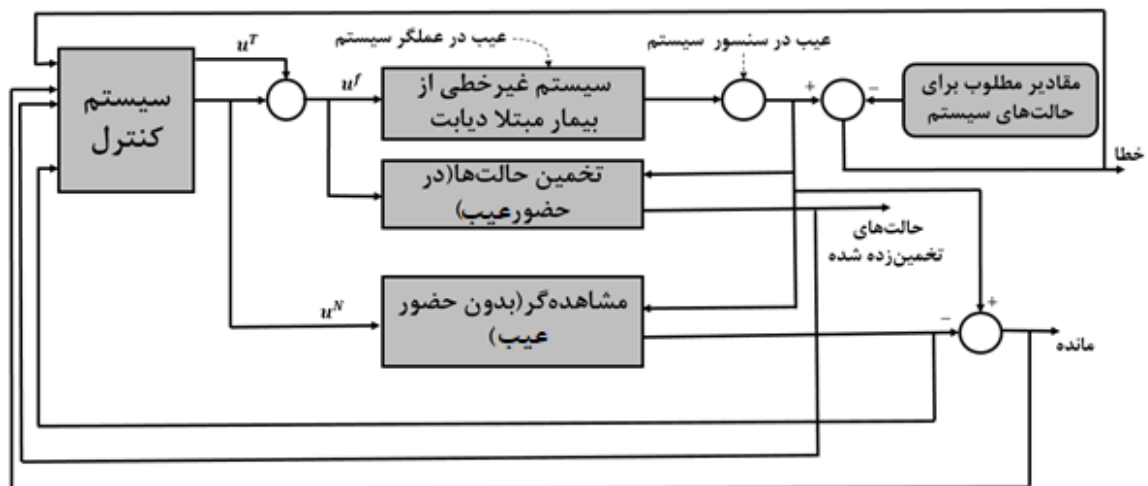
شایستگی و کنترل مدلغزشی

۶-۱- مقدمه

در این فصل دومین نوآوری رساله مطرح گردیده است. ایده اساسی این کار الهام گرفته شده از کار [۳۴] می‌باشد. که در آن مقاله در دو حالت استفاده از (شبکه عصبی و کنترل مدلغزشی) و (کنترل مدلغزشی و الگوریتم بازیگر-نقاد) کنترل تحمل‌پذیر عیب انجام شده است. کنترل کننده اصلی کنترل کننده مدلغزشی می‌باشد و کنترل کننده کمکی که بر پایه یادگیری تقویتی طراحی می‌شود وظیفه کنترل عیب را بر عهده دارد. در کنترلرلهایی که بر پایه یادگیری تقویتی طراحی گردیده‌اند از دو الگوریتم بازیگر-نقاد و مسیرهای شایستگی استفاده شده است.

۶-۲- ساختار کلی

ساختار کلی فرآیند کنترلی در شکل ۶-۱ نشان داده شده است. برای مشاهده‌گر و تخمین‌گر حالت‌ها از شبکه عصبی توابع پایه شعاعی استفاده شده است [۳۴]. مشاهده‌گر، سیستم بیمار دیابتی در عدم حضور عیب را شناسایی می‌کند و تخمین‌گر حالت‌ها وظیفه تخمین حالت‌های سیستم در زمان رخداد عیب بر عهده دارد. کل سیستم ذکر شده در این فصل نیز بر اساس نمونه‌برداری از داده‌ها کنترل تحمل‌پذیر عیب را انجام می‌دهد.



شکل ۶-۱: سیستم کنترل تحمل‌پذیر عیب با ترکیب کنترل مد‌لغزشی و شبکه‌های عصبی مصنوعی

هدف ما در این قسمت استفاده از یکی از روش‌های نوین یادگیری تقویتی به نام مسیر شایستگی و ترکیب آن با کنترل مد لغزشی جهت کنترل تحمل‌پذیر عیب می‌باشد و در نهایت روش پیشنهادی با روش‌های قبلی مقایسه گردیده است که نتایج شبیه‌سازی آن در فصل بعد آورده شده است:

با توجه به شکل (۶-۱)، ورودی کنترلی u^f میزان دوز داروی تزریقی به بدن بیمار برای تنظیم میزان گلوکز می‌باشد. این میزان دوز دارو در هر گام به‌گونه‌ای تعیین می‌شود که در حضور عیب در عملگر و یا سنسور سیستم باز هم قادر به کنترل میزان گلوکز خون و نگهداشتن آن بر روی مقدار مطلوب باشد.

در این رساله سطح لغزش بر روی مقادیر خطا و مانده تعریف شده است. به صورت کلی روابط آن در معادلات (۶-۱) و (۶-۲) بیان شده اند.

$$s^N = \binom{n-1}{0} e_n^N + \binom{n-1}{1} \lambda^1 e_{n-1}^N + \dots + \binom{n-1}{n-3} \lambda^{n-3} e_3^N + \binom{n-1}{n-2} \lambda^{n-2} e_2^N + \binom{n-1}{n-1} \lambda^{n-1} e_1^N \quad (6-1)$$

در رابطه (۶-۱)، منظور از حرف N سیستم نرمال می‌باشد. s^N تعریف سطح لغزش بر روی سیستم نرمال و s^R تعریف سطح لغزش بر روی مقدار مانده می‌باشد. در رابطه (۶-۲)، e اختلاف میان مقدار مطلوب و مقدار خروجی سیستم اصلی در هر لحظه می‌باشد. e_i ها نیز نشان دهنده خطای مابین حالت‌های سیستم واقعی (در حضور یا عدم حضور عیب) و مقدار مطلوب حالت‌ها می‌باشد. λ نیز همواره مقداری ثابت مثبت به خود می‌گیرد.

$$s^R = \binom{n-1}{0} R_n + \binom{n-1}{1} \lambda^1 R_{n-1} + \dots + \binom{n-1}{n-3} \lambda^{n-3} R_3 + \binom{n-1}{n-2} \lambda^{n-2} R_2 + \binom{n-1}{n-1} \lambda^{n-1} R_1 \quad (6-2)$$

در رابطه (۶-۲)، R مقدار مانده بوده و اختلاف خروجی سیستم اصلی و سیستم تخمین زده شده می‌باشد.

فرمول محاسبه خطا براساس اختلاف مقدار خروجی اصلی و مقدار مطلوب در رابطه (۶-۳) نمایش داده شده است.

$$e = y_{new} - x_d \quad (6-3)$$

در رابطه (۶-۳)، y_{new} میزان خروجی در لحظه فعلی می‌باشد. x_d مقدار مطلوب خروجی می‌باشد. مقدار مانده طبق رابطه (۶-۴) به دست می‌آید که اختلاف میزان خروجی اصلی و مقدار تخمین زده شده از سیستم اصلی می‌باشد.

$$R = y_{new} - \hat{y}_n \quad (4-6)$$

در رابطه (۶-۴)، $\hat{y}_n$ مقدار تخمین زده شده از خروجی سیستم اصلی می‌باشد. تابع لیاپانوف تعریف شده در این مقاله برای اثبات پایداری سیستم به صورت معادله (۶-۵) تعریف شده است [۳۴].

$$v(t) = \frac{1}{2} (s^R)^2 \quad (6-5)$$

برای برقراری پایداری سیستم شرط بیان شده در رابطه (۶-۶) باید برقرار باشد.

$$\dot{v} = s^R \dot{s}^R \leq 0 \quad (6-6)$$

در اینجا برای پایداری سیستم در هر لحظه تابع فوق که بیان کننده تابع انرژی سیستم می باشد، در تابع پاداش بیان شده است و با کمتر شدن هر چه بیشتر آن عامل پاداش بیشتری از محیط دریافت می کند. برای تخمین سیستم نرمال در حالتی که عیب در سیستم وجود ندارد از شبکه عصبی توابع پایه شعاعی استفاده شده است. تخمین نرمال یک سیستم مرتبه n در رابطه (6-7) بیان شده است [34].

$$\begin{cases} \dot{\hat{x}}_1^N = \hat{x}_2^N \\ \dot{\hat{x}}_2^N = \hat{x}_3^N \\ \vdots \\ \dot{\hat{x}}_k^N = \hat{x}_{k+1}^N \\ \vdots \\ \dot{\hat{x}}_n^N = f_{est.}(\hat{x}^N) + u^N \\ \hat{y}^N = \hat{x}_1^N \end{cases} \quad (6-7)$$

در رابطه (6-7)، u^N ورودی کنترلی برای سیستم نرمال بوده و مقدار آن در هر لحظه با توجه به رابطه (6-8) محاسبه می شود.

$$u^N = \eta_N \operatorname{sgn}(s^N) - \binom{n-1}{0} (\hat{f}_N(\hat{X}^N) - x_d^{(n)}) - \binom{n-1}{1} \lambda^1 e_n^N - \dots - \binom{n-1}{n-2} \lambda^{n-2} e_3^N - \binom{n-1}{n-1} \lambda^{n-1} e_2^N \quad (6-8)$$

در رابطه (6-8)، η_N مقداری ثابت و مثبت می باشد. مقدار این پارامتر که عددی بین صفر و یک می باشد در شبیه سازی ها با سعی و خطا تعیین گردیده است. $\hat{f}_N(\hat{X}^N)$ تخمین حالت ها در سیستم نرمال در هر لحظه می باشد. e ، اختلاف میان خروجی سیستم اصلی و خروجی تخمین زده شده با استفاده از شبکه عصبی می باشد.

در حالتی که سیستم دچار عیب شده است، حالت‌های سیستم با استفاده از روش شبکه‌های عصبی توابع پایه شعاعی مشاهده شده و با خروجی سیستم اصلی مقایسه می‌شوند. ورودی در این حالت u^f می‌باشد که طبق رابطه (۹-۶) محاسبه می‌شود. که ورودی کنترلی سیستم اصلی نیز می‌باشد.

$$u^f = u^N + u^T \quad (۹-۶)$$

در رابطه (۹-۶)، u^T ورودی کنترلی تحمل‌پذیر^۱ بوده و بر اساس رابطه (۱۰-۶) محاسبه می‌شود.

$$u^T = \eta_R \operatorname{sgn}(s^R) - \binom{n-1}{0} (\hat{f}_f(\hat{X}^N) - \hat{f}_N(\hat{X}^N)) - \binom{n-1}{1} \lambda^1 R_n - \dots - \binom{n-1}{n-2} \lambda^{n-2} R_3 - \binom{n-1}{n-1} \lambda^{n-1} R_2 \quad (۱۰-۶)$$

در رابطه (۱۰-۶)، η_R مقداری ثابت و مثبت می‌باشد. $\hat{f}_f(\hat{X}^N)$ تخمین حالت‌های سیستم در شرایطی که سیستم دچار عیب شده باشد و $\hat{f}_N(\hat{X}^N)$ تخمین حالت‌های سیستم در شرایط نرمال می‌باشد. برای اثبات پایداری سیستم تابع لیاپانوف به صورت معادله (۱۱-۶) تعریف می‌گردد.

$$V_T = \frac{1}{2} (s^N)^2 + \frac{1}{2} (s^R)^2 + \frac{1}{2} \tilde{W}_n^T \tilde{W}_n + \frac{1}{2} \tilde{W}_f^T \tilde{W}_f \quad (۱۱-۶)$$

در رابطه فوق، $\tilde{W}_n$ بردار خطای وزن‌دهی شده برای سیستم نرمال و $\tilde{W}_f$ بردار خطای وزن‌دهی شده برای سیستم دارای عیب می‌باشد. با مشتق‌گیری از طرفین رابطه فوق رابطه (۱۲-۶) بدست می‌آید.

$$\dot{V}_T = s^N \dot{s}^N + s^R \dot{s}^R + \tilde{W}_n^T \dot{\tilde{W}}_n + \tilde{W}_f^T \dot{\tilde{W}}_f \quad (۱۲-۶)$$

در رابطه فوق، $\dot{s}^N$ و $\dot{s}^R$ بر اساس روابط (۱۳-۶) و (۱۴-۶) به‌دست می‌آیند.

$$\dot{s}^N = \binom{n-1}{0} \dot{e}_n^N + \binom{n-1}{1} \lambda^1 e_n^N + \dots + \binom{n-1}{n-3} \lambda^{n-3} e_4^N + \binom{n-1}{n-2} \lambda^{n-2} e_3^N + \binom{n-1}{n-1} \lambda^{n-1} e_2^N \quad (۱۳-۶)$$

$$\dot{s}^R = \binom{n-1}{0} \dot{R}_n + \binom{n-1}{0} \lambda^1 R_n + \dots + \binom{n-1}{n-3} \lambda^{n-3} R_4 + \binom{n-1}{n-2} \lambda^{n-2} R_3 + \binom{n-1}{n-1} \lambda^{n-1} R_2 \quad (۱۴-۶)$$

^۱ Tolerant Control Input

در روابط فوق $\dot{e}_n^N = (f(X^N) - x_d^{(n)} + u^N)$ و $\dot{R}_n = (f(X^f) - f(X^N))$ را جایگزین کرده و روابط (۱۵-۶) و (۱۶-۶) بدست می‌آیند.

$$\begin{aligned} \dot{s}^N = & \left(n-1 \atop 0\right) (f(X^N) - x_d^{(n)} + u^N) + \left(n-1 \atop 1\right) \lambda^1 e_n^N + \dots \\ & + \left(n-1 \atop n-3\right) \lambda^{n-3} e_4^N + \left(n-1 \atop n-2\right) \lambda^{n-2} e_3^N + \left(n-1 \atop n-1\right) \lambda^{n-1} e_2^N \end{aligned} \quad (۱۵-۶)$$

$$\begin{aligned} \dot{s}^R = & \left(n-1 \atop 0\right) (f(X^f) - f(X^N)) + \left(n-1 \atop 0\right) (F_a + F_s^{(n)} + d + u^f + u^N) \\ & + \left(n-1 \atop 1\right) \lambda^1 R_n + \dots + \left(n-1 \atop n-3\right) \lambda^{n-3} R_4 + \left(n-1 \atop n-2\right) \lambda^{n-2} R_3 \\ & + \left(n-1 \atop n-1\right) \lambda^{n-1} R_2 \end{aligned} \quad (۱۶-۶)$$

در رابطه (۱۶-۶) u^f ورودی سیستم دارای عیب می‌باشد که طبق رابطه (۹-۶) محاسبه می‌شود. که ورودی کنترلی سیستم اصلی نیز می‌باشد. با جایگذاری داریم:

با جایگزینی روابط فوق در رابطه (۱۲-۶) $\dot{V}_T$ بدست می‌آید. در رابطه (۱۷-۶) قابل نمایش می‌باشد.

$$\begin{aligned} \dot{V}_T = & -\eta_N |s^N| - \eta_R |s^R| + \tilde{w}_n^T \dot{\tilde{w}}_n + \tilde{w}_f^T \dot{\tilde{w}}_f + (s^N - s^R) \left(f(X^N) - \right. \\ & \left. \hat{f}_N(\hat{X}^N) \right) + s^R \left(n-1 \atop 0\right) (f(X^f) - \hat{f}_f(\hat{X}^N)) + s^R \left(n-1 \atop 0\right) (F_a + F_s^{(n)} + d) \end{aligned} \quad (۱۷-۶)$$

در روابط فوق:

$$f(X^N) - \hat{f}_N(\hat{X}^N) = \tilde{w}_n^T G_n(0) \quad (۱۸-۶)$$

$$f(X^f) - \hat{f}_f(\hat{X}^N) = \tilde{w}_f^T G_f(0) \quad (۱۹-۶)$$

G_f و G_n تخمین‌گر حالت‌ها برای سیستم نرمال و سیستم دارای عیب می‌باشد. با جایگزینی روابط

فوق در رابطه (۱۷-۶) و ساده سازی، $\dot{V}_T$ به فرم رابطه (۲۰-۶) خواهد بود.

$$\dot{V}_T = -\eta_N |s^N| - \eta_R |s^R| < 0 \quad (۲۰-۶)$$

اثبات رابطه (۲۰-۶) بر اساس روابط زیر خواهد بود.

$$V_T = V_N + V_R \quad (۲۱-۶)$$

در رابطه (۲۲-۶) V_N به صورت رابطه (۲۲-۶) می باشد.

$$V_N = \frac{1}{2}(s^N)^2 + \frac{1}{2}\tilde{w}_n^T \tilde{w}_n \quad (۲۲-۶)$$

و V_R نیز بر اساس رابطه (۲۳-۶) می باشد.

$$V_R = \frac{1}{2}(s^R)^2 \quad (۲۳-۶)$$

با مشتق گیری از رابطه (۲۳-۶) رابطه (۲۴-۶) بدست می آید.

$$\begin{aligned} \dot{V}_N = & s^N \binom{n-1}{0} (f(X^N) - x_d^{(n)} + u^N) + s^N \binom{n-1}{1} \lambda^1 e_n^N + \dots + \\ & s^N \binom{n-1}{n-3} \lambda^{n-3} e_4^N + s^N \binom{n-1}{n-2} \lambda^{n-2} e_3^N + s^N \binom{n-1}{n-1} \lambda^{n-1} e_2^N + \tilde{w}_n^T \dot{\tilde{w}}_n \end{aligned} \quad (۲۴-۶)$$

با ساده سازی رابطه فوق، رابطه (۲۴-۶) حاصل می شود.

$$\dot{V}_N = -\eta_N |s^N| + s^N (f(X^N) - \hat{f}_N(\hat{X}^N)) + \tilde{w}_n^T \dot{\tilde{w}}_n \quad (۲۴-۶)$$

با در نظر گرفتن $\dot{\tilde{w}}_n = -\tilde{w}_n = s^N G_n(0)$ رابطه (۲۵-۶) برقرار می باشد.

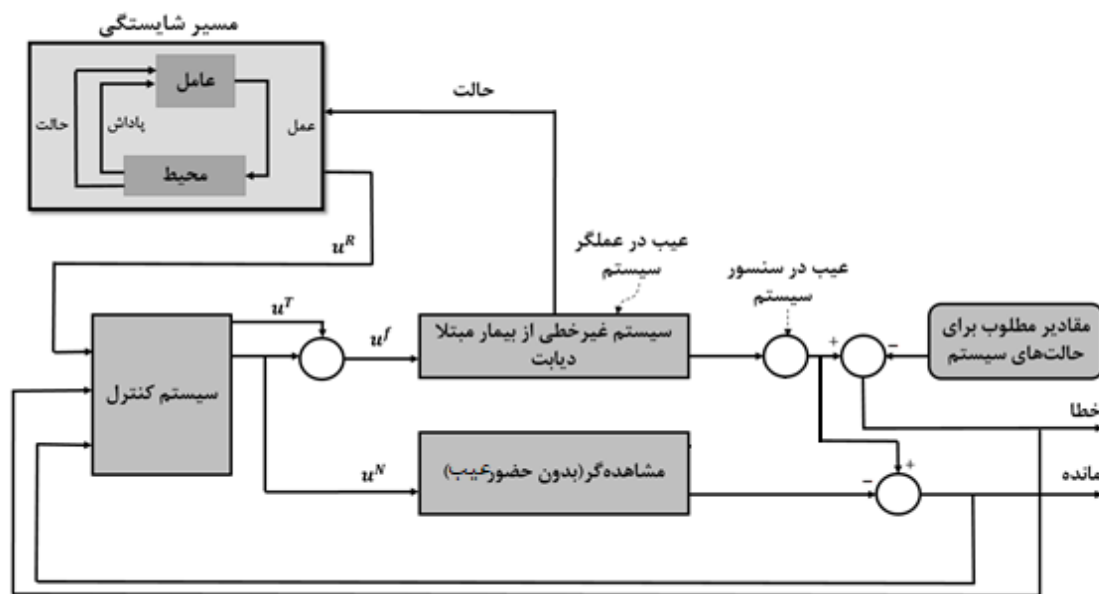
$$\dot{V}_N = -\eta_N |s^N| < 0 \quad (۲۵-۶)$$

همچنین با مشتق گیری از رابطه (۲۳-۶)، رابطه (۲۶-۶) حاصل می شود

$$\dot{V}_R = s^R \dot{s}^R \quad (۲۶-۶)$$

۳-۶ - ساختار پیشنهادی

اکنون همان ساختار استفاده شده را در این رساله استفاده کرده ایم منتهی به جای آنکه از شبکه عصبی استفاده کنیم از مسیرهای شایستگی که یکی از الگوریتم های نوین و کارآمد یادگیری تقویتی می باشد استفاده کرده ایم. ساختار کلی پیشنهادی به صورت زیر می باشد:



شکل ۶-۲: کنترل میزان گلوکز با استفاده از روش مسیرهای شایستگی در زمان رخداد عیب

در بخش مربوط به یادگیری تقویتی، تابع پاداش با توجه به رابطه (۶-۲۷) تعیین می‌شود.

$$r(t) = -\lambda_R (s^R \dot{s}^R) \quad (۶-۲۷)$$

در رابطه (۶-۲۷)، λ_R مقداری ثابت و مثبت می‌باشد. $\dot{s}^R$ مشتق تابع s^R بوده و طبق رابطه (۶-۲۸) محاسبه می‌شود.

$$\begin{aligned} \dot{s}^R = & \begin{pmatrix} 2 \\ 0 \end{pmatrix} (f(x^f) - f(x^N)) + \begin{pmatrix} 2 \\ 0 \end{pmatrix} (F_a + F_s^{(n)} + d + u^f + u^N) + \begin{pmatrix} 2 \\ 1 \end{pmatrix} \lambda^1 R_3 \\ & + \begin{pmatrix} 2 \\ 2 \end{pmatrix} \lambda^2 R_2 \end{aligned} \quad (۶-۲۸)$$

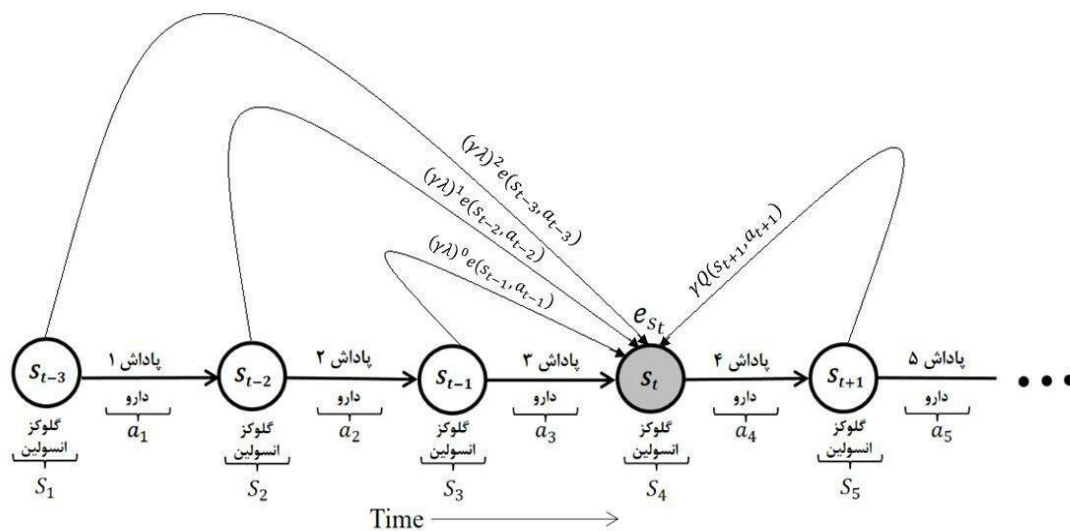
با توجه به رابطه (۶-۲۷)، زمانی که $s^R \dot{s}^R$ مثبت باشد، انرژی سیستم افزایش پیدا کرده و پایداری کم می‌شود، بنابراین پاداش منفی خواهد بود. اما در غیر این صورت پایداری افزایش یافته و پاداش دریافتی مثبت می‌شود. پس رابطه (۶-۲۹) برقرار خواهد بود.

$$\dot{V}_T = (\dot{V}_N + \dot{V}_R) < 0 \quad (۶-۲۹)$$

با توجه به رابطه فوق میزان انرژی کاهش پیدا کرده و پایداری سیستم تضمین می‌گردد. ورودی کنترلی u^T نیز بر اساس عملی که عامل انتخاب می‌کند محاسبه شده و طبق رابطه (۶-۳۰) بدست می‌آید.

$$u^T = -u^R - (\eta_R \text{sgn}(s^R)) \quad (۶-۳۰)$$

در رابطه (۶-۳۰)، η_R مقداری ثابت و مثبت می‌باشد. مقدار این پارامتر که عددی بین صفر و یک می‌باشد در شبیه‌سازی‌ها با سعی و خطا تعیین گردیده است. u^R عملی می‌باشد که توسط عامل انتخاب شده است. همانطور که اشاره شده در این بخش برای کنترل تحمل‌پذیر عیب از روش مسیرهای شایستگی استفاده شده است. شکل ۶-۳ شمای کلی از نحوه به‌روزرسانی ارزش حالت-عمل در الگوریتم مسیرهای شایستگی در کنترل گلوکز خون را نشان می‌دهد.



شکل ۶-۳: شمای کلی از نحوه به‌روزرسانی ارزش حالت-عمل در روش مسیرهای شایستگی

کنترل میزان گلوکز با استفاده از روش مسیرهای شایستگی و تطبیق آن با ساختار فوق در شبه کد شکل ۶-۴ آورده شده است.

شروع
(۱) دریافت شرایط اولیه سیستم
(۲) تنظیم مقادیر اولیه برای پارامترهای ثابت $\alpha, \gamma, \lambda, \dots$
(۳) تعیین حالت‌ها و عمل‌ها و تنظیم جدول Q بر اساس آن‌ها
(۴) تا زمانی که میزان گلوکز در بدن بیمار به مقدار مطلوب نرسیده است، مراحل زیر را تکرار کن:
(۱,۴) مشاهده میزان گلوکز، مانده و خطای بین میزان گلوکز محاسبه شده در هر گام و مقدار مرجع گلوکز
(۲,۴) انتخاب عمل a_t با استفاده از سیاست $\varepsilon - greedy$ (u^R)
(۳,۴) اعمال دارو بر روی مدل غیر خطی از بیمار مبتلا به دیابت (u^f)
(۴,۴) محاسبه پاداش دریافتی در حالت فعلی (s_t)
(۵) بروز رسانی جدول Q بر اساس روش مسیرهای شایستگی
(۱,۵) ساخت ماتریس شایستگی (e) بر اساس ابعاد ماتریس Q
(۲,۵) محاسبه مقدار δ بر اساس رابطه:
$\delta = r + \gamma Q(s', a) - Q(s, a)$
(۳,۵) محاسبه مقدار شایستگی بر اساس رابطه زیر:
$e_t(s, a) = e_t(s, a) + 1$
(۴,۵) بروز رسانی جدول Q با استفاده از رابطه زیر:
$Q(s, a) = Q(s, a) + \alpha \delta e(s, a)$
(۵,۵) بروز رسانی مقدار شایستگی با استفاده از رابطه زیر:
$e_t(s, a) = \gamma \lambda \times e_t(s, a)$
(۶) در نظر گرفتن شرایط اولیه جدید یا قرار گرفتن در حالت s_{t+1}

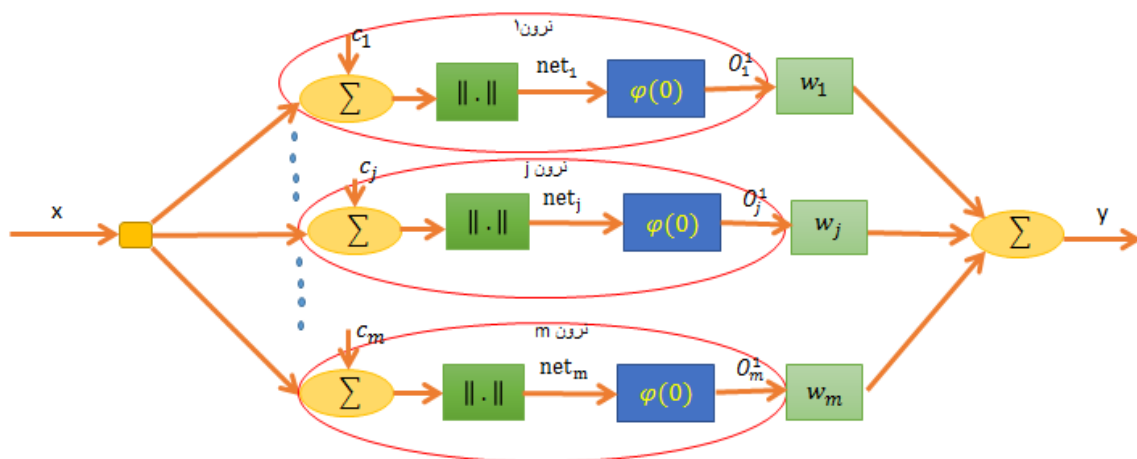
شکل ۴-۶: کنترل میزان گلوکز با استفاده از روش مسیرهای شایستگی

در قسمت شناسایی سیستم در طرح مذکور از شبکه‌های عصبی توابع پایه شعاعی استفاده شده است که به اختصار در اینجا توضیح داده می‌شود.

۶-۳-۱- توابع پایه شعاعی^۱

توابع پایه شعاعی میزان شباهت بین بردار ورودی و بردار نمونه را بررسی می‌کند. یک نرون برای خروجی وجود دارد. لایه خروجی یک تابع خطی می‌باشد. تعداد نرون‌های لایه مخفی بیشتر از تعداد ورودی‌ها است. کاربرد این شبکه‌ها شامل تقریب توابع، درونیابی، سری‌های زمانی و کلاس‌بندی می‌باشد [۹۶].

¹ Radial Basis Function



شکل ۵-۶: نمایش توابع پایه شعاعی (نمای داخلی شبکه)

همانطور که در شکل (۵-۶) مشاهده می‌شود، یک بردار ورودی وجود دارد و این بردار ورودی به تمام نرون‌های لایه مخفی می‌رود.

c_i بردار نمونه‌برداری می‌باشد و برای هر نرون به طور جداگانه می‌باشد. مشابه بایاس می‌باشد و وزن خود را از جای دیگری نمی‌گیرد [۹۶]. از رابطه (۳۱-۶) برای تعیین میزان شباهت استفاده می‌شود.

$$net_j = \| (X - c_j) \| \quad (۳۱-۶)$$

$$net_j = \sqrt{(X_1 - c_{1j})^2 + (X_2 - c_{2j})^2 + \dots + (X_n - c_{nj})^2} \quad (۳۲-۶)$$

$$O_j^1 = \theta(net_j) \quad (۳۳-۶)$$

در رابطه (۳۳-۶)، θ تابع فعال‌ساز می‌باشد که خروجی هر نرون وارد آن می‌شود. معروف‌ترین تابع فعال‌ساز در این شبکه تابع فعال‌ساز گوسی می‌باشد که معرفی این تابع در رابطه (۳۴-۶) آمده است.

$$\theta(net_j) = e^{-\frac{1}{2}(\frac{net_j}{\sigma_j})^2} \quad (۳۴-۶)$$

در رابطه فوق، σ_j میزان انحراف معیار از مرکز دسته می‌باشد.

پارامترهای شبکه:

۱- ورودی:

$$X = [x_1, x_2, x_3, \dots, x_n] \quad (۳۵-۶)$$

۲- تعریف بردار نمونه ها (C):

$$c_j = [c_{1j}, c_{2j}, c_{3j}, \dots, c_{nj}] \quad (۳۶-۶)$$

۳- خروجی برای هر لایه مخفی (m تا نورون وجود دارد)

$$O^1 = [O_1^1, O_2^1, O_3^1, \dots, O_m^1]' \quad (۳۷-۶)$$

۴- انحراف معیار برای هر نورون:

$$\sigma = [\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_m] \quad (۳۸-۶)$$

۵- وزن های اتصال از لایه مخفی و لایه خروجی:

$$w = [w_1, w_2, w_3, \dots, w_m]' \quad (۳۹-۶)$$

طرح پیشنهادی برای کنترل تحمل پذیر خطا با استفاده از مسیرهای شایستگی برای دو مدل بیماری دیابت شبیه سازی شده است.

۴-۶- شبیه سازی ساختارهای کنترلی فوق بر روی مدل دان کومار و سندیا

مدل ریاضی که در این قسمت شبیه سازی روش های فوق بر روی آن پیاده سازی شده است توسط دان کومار و سندیا^۱ ارائه شده است [۸۲]. آن ها دینامیک بدن بیمار مبتلا به دیابت را با سه معادله دیفرانسیل توصیف کرده اند. این معادلات در روابط (۴-۱) تا (۴-۳) بیان شدند.

این مدل مرتبه ۳ شامل متغیرهای G، گلوکز فرد، X، متغیر عمومی انسولین برای جز گذشته و I انسولین کنونی فرد می باشد. متغیر انسولین قبلی، میزان برداشت انسولین را بین گلوکز خون و بافت های محیطی و کبد تنظیم می کند. از مزیت های این مدل می توان به وجود ثوابت Gb و Ib اشاره کرد که نشان دهنده مقادیر پایه ای برای گلوکز و انسولین می باشد. متغیر u در رابطه (۴-۴)، میزان دوز داروی تزریقی به بیمار برای کنترل میزان گلوکز می باشد. ثوابت سیستم به همراه مقادیری که در شبیه سازی ها برای فرد در نظر گرفته شده اند در جدول (۴-۱) آورده شده اند.

¹ Sandhya danKumar

جدول ۶-۱: ثوابت سیستم [۸۲]

پارامتر	توضیحات	واحد	مقدار
G_b	مقدار پایه‌ای گلوکز پلاسمایی قبل از تزریق	$\frac{mg}{dl}$	۸۰
I_b	مقدار پایه‌ای انسولین پلاسمایی قبل از تزریق	$\frac{\mu U}{ml}$	۷
m_1	میزان بازجذب گلوکز در عضلات و کبد و بافت‌های چربی که مستقل از سطح انسولین می‌باشد.	min^{-1}	۰,۰۳۱۷
m_2	میزان کاهش در نرخ بازجذب گلوکز توسط بافت‌ها	min^{-1}	۰,۰۱۷
m_3	میزان افزایش در نرخ بازجذب گلوکز در بافت‌ها (که مستقل از میزان گلوکز می‌باشد) به ازای هر واحد از غلظت انسولین در بالای I_b	$min^{-2}(\frac{\mu U}{ml})$	5.3×10^{-6}
m_4	میزان رهایش انسولین از سلول‌های B پانکراس، بعد از تزریق گلوکز در حالی که غلظت گلوکز بالای h می‌باشد	$\frac{\mu U}{ml} min^{-2}(\frac{mg}{dl})^{-1}$	۰,۰۰۴۲
m_5	میزان سطح آستانه‌ای از گلوکز که بالای آن مقدار سلول‌های B پانکراس انسولین ترشح می‌کنند	---	۸۰,۲۵
m_6	نرخ تخریب درجه اول برای انسولین در پلازما	min^{-1}	۰,۲۶۴

۶-۴-۱- کنترل میزان قند خون با استفاده از شبکه‌های عصبی

برای سیستم بیان شده در این مقاله تعریف سطوح لغزش به شکل معادلات (۶-۴۰) و (۶-۴۱) خواهد بود.

$$s^N = \binom{2}{0} e_3^N + \binom{2}{1} \lambda_c^1 e_2^N + \binom{2}{2} \lambda_c^2 e_1^N \quad (6-40)$$

$$s^R = \binom{2}{0} Re_3 + \binom{2}{1} \lambda_c^1 Re_2 + \binom{2}{2} \lambda_c^2 Re_1 \quad (6-41)$$

ورودی کنترلی u^N برای سیستم استفاده شده در این مقاله در معادله (۶-۴۲) بیان شده است.

$$u^N = \eta_N sgn(s^N) - \binom{2}{0} (\hat{f}_N(\hat{X}^N) - x_d^{(n)}) - \binom{2}{1} \lambda_c^1 e_3^N - \binom{2}{2} \lambda_c^2 e_2^N \quad (6-42)$$

برای محاسبه میزان دوز داروی تزریقی به بیمار از رابطه (۶-۴۳) استفاده می‌شود.

$$u^f = u^N + u^T \quad (6-43)$$

که ورودی کنترل تحمل‌پذیر از رابطه (۶-۴۴) به‌دست می‌آید

$$u^T = \eta_R \text{sgn}(s^R) - \binom{2}{0} (\hat{f}_f(\hat{X}^N) - \hat{f}_N(\hat{X}^N)) - \binom{2}{1} \lambda_c^1 Re_3 - \binom{2}{2} \lambda_c^2 Re_2 \quad (6-44)$$

۶-۴-۲ کنترل میزان قند خون با استفاده از مسیرهای شایستگی

در کنترل میزان گلوکز خون با استفاده از روش مسیرهای شایستگی تعریف سطوح لغزش برای سیستم نرمال و مقدار مانده مطابق با موارد بیان شده در بخش (۶-۳-۱) می‌باشد.

در بخش مربوط به یادگیری تقویتی، حالت‌های در نظر گرفته شده دوبعدی بوده و میزان گلوکز خون و انسولین می‌باشد. تابع پاداش با توجه به رابطه (۶-۴۵) تعیین می‌شود

$$r(t) = -\lambda_R (s^R \dot{s}^R) \quad (6-45)$$

در رابطه (۶-۴۵)، λ_R مقداری ثابت و مثبت می‌باشد. $\dot{s}^R$ مشتق تابع s^R بوده و طبق رابطه (۶-۴۶) محاسبه می‌شود.

$$\dot{s}^R = \binom{2}{0} (f(x^f) - f(x^N)) + \binom{2}{0} (F_a + F_s^{(n)} + d + u^f + u^N) + \binom{2}{1} \lambda^1 Re_3 + \binom{2}{2} \lambda^2 Re_2 \quad (6-46)$$

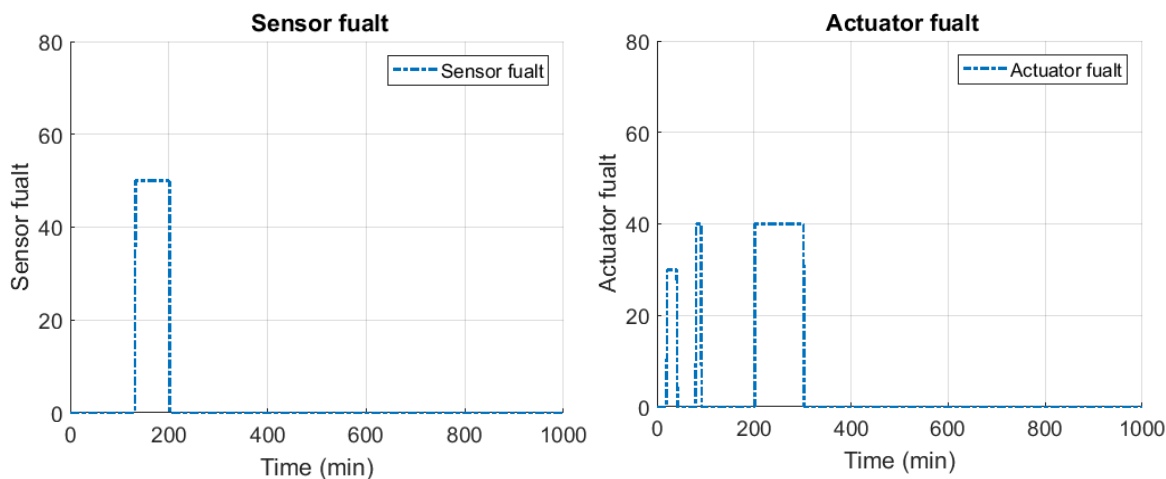
ورودی کنترلی u^T نیز بر اساس عملی که عامل انتخاب می‌کند محاسبه شده و طبق رابطه (۶-۴۷) بدست می‌آید.

$$u^T = -u^R - (\eta_R \text{sgn}(s^R)) \quad (6-47)$$

در رابطه (۶-۴۷)، η_R مقداری ثابت و مثبت می‌باشد. u^R عملی می‌باشد که توسط عامل انتخاب شده است. کنترل میزان گلوکز با استفاده از روش مسیرهای شایستگی در شکل (۶-۳) آورده شده است.

شرایط اولیه که برای بیمار در نظر گرفته شده است، بدین ترتیب می‌باشد: $I_0 = 25$, $X_0 = 10$ and $G_0 = 250$. مقادیر مطلوب برای متغیرها که در نهایت می‌بایست به آن برسند، برای گلوکز ۸۵، مقدار عمومی انسولین ۱۸ و انسولین خون فرد ۵۰ می‌باشد. دوز داروی تزریقی که متغیر u^f می‌باشد، باید

به گونه‌ای تعیین شود که هنگام بروز عیب در سیستم که در سنسور و عملگر اتفاق می‌افتد، قادر به کنترل میزان گلوکز و نگه‌داشتن آن در سطح مطلوب باشد عیوب وارد شده به عملگر و سنسور سیستم به ترتیب در شکل‌های (۶-۷) و (۶-۸) نمایش داده شده است.



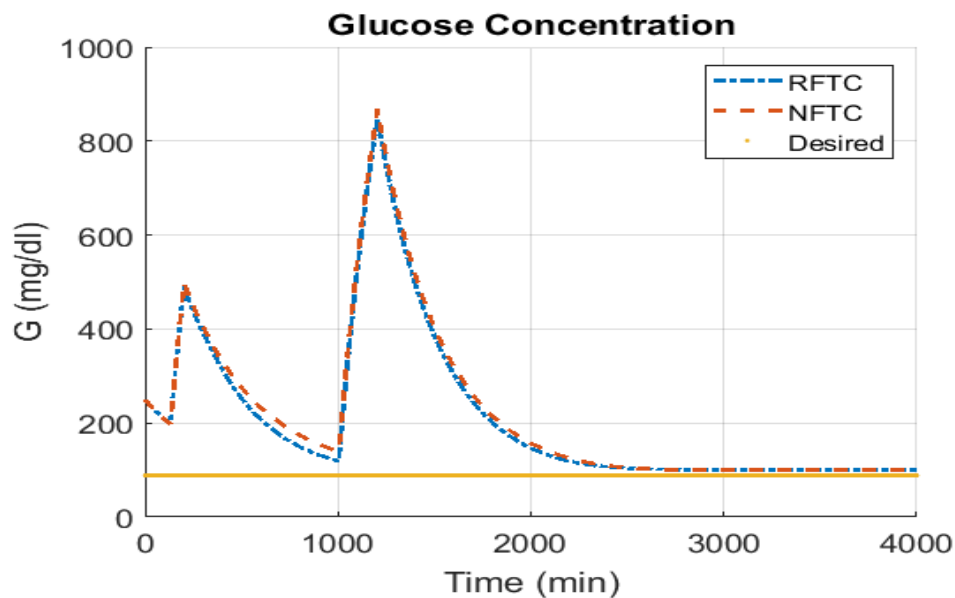
شکل ۶-۷: عیب وارد شده به سنسور سیستم

شکل ۶-۶: عیب وارد شده به عملگر سیستم

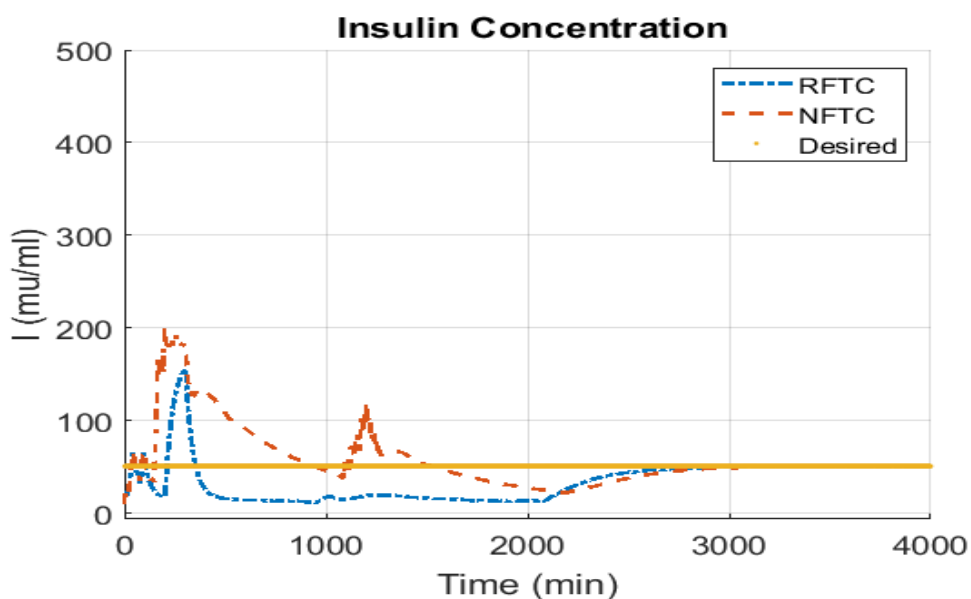
نتایج حاصل از شبیه‌سازی برای میزان گلوکز، انسولین کنونی فرد و دوز داروی تزریقی رسم شده است. شایان ذکر است که در ادامه تصاویر هرجا عنوان ^۱ (RFTC) آورده شده است به معنای استفاده از روش مسیرهای شایستگی در کنترل تحمل‌پذیر خطا می‌باشد و هرجا عنوان ^۲ (NFTC) آورده شده است به معنای استفاده از روش مسیرهای شایستگی در کنترل تحمل‌پذیر خطا می‌باشد. شکل (۶-۹) کنترل میزان گلوکز را در یک بیمار مبتلا به دیابت در حضور هر دو عیب سنسور و عملگر را نمایش می‌دهد. شکل (۶-۱۰) نمایش‌گر میزان انسولین برای کنترل میزان گلوکز با استفاده از هر دو روش می‌باشد.

^۱ RL Fault Tolerant Control

^۲ Neural network Fault Tolerant Control



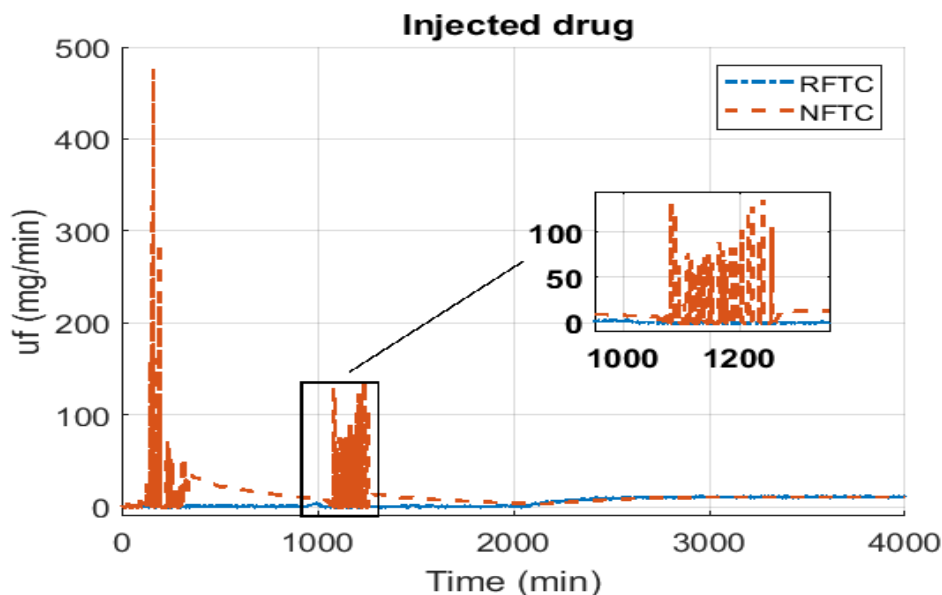
شکل ۶-۸: میزان گلوکز در یک بیمار مبتلا به دیابت در حضور هر دو عیب سنسور و عملگر



شکل ۶-۹: میزان انسولین در یک بیمار مبتلا به دیابت در حضور هر دو عیب سنسور و عملگر

همان‌طور که در شکل (۶-۹) مشاهده می‌شود، هر دو روش قادر به کنترل میزان گلوکز و نگاه‌داشتن آن روی مقدار مطلوب می‌باشند، اما روش یادگیری تقویتی با سرعت بیشتر توانسته است میزان گلوکز را کنترل کرده و به مقدار مطلوب برساند. شکل (۶-۱۰) نمایش‌گر میزان انسولین برای کنترل میزان گلوکز با استفاده از هر دو روش می‌باشد. با توجه به شکل (۶-۱۰)، میزان انسولین توسط روش

یادگیری تقویتی نسبت به روش شبکه‌های عصبی بسیار کمتر می‌باشد. با وجود اینکه در بعضی مواقع میزان انسولین کمتر از مقدار مطلوب می‌باشد، اما باز هم روش‌های انتخابی قادر به کنترل و رساندن آن به مقدار مطلوب بوده‌اند. میزان دوز داروی تزریق شده برای کنترل گلوکز با استفاده از هر دو روش در شکل (۱۱-۶) قابل مشاهده می‌باشد.



شکل ۶-۱۰: میزان دوز داروی تزریقی به بیمار با استفاده از روش یادگیری تقویت و شبکه‌های عصبی در هنگام رخداد عیب
همانطور که در شکل (۱۱-۶) نشان داده شده است، میزان داروی تزریقی به بیمار با استفاده از روش یادگیری تقویتی نسبت به روش شبکه‌های عصبی کمتر می‌باشد. مجموع دوز داروی تزریقی با استفاده از هر دو روش در جدول (۶-۲) بیان شده است.

جدول ۶-۲: مجموع دوز داروی تزریقی با استفاده از روش شبکه‌های عصبی و یادگیری تقویتی

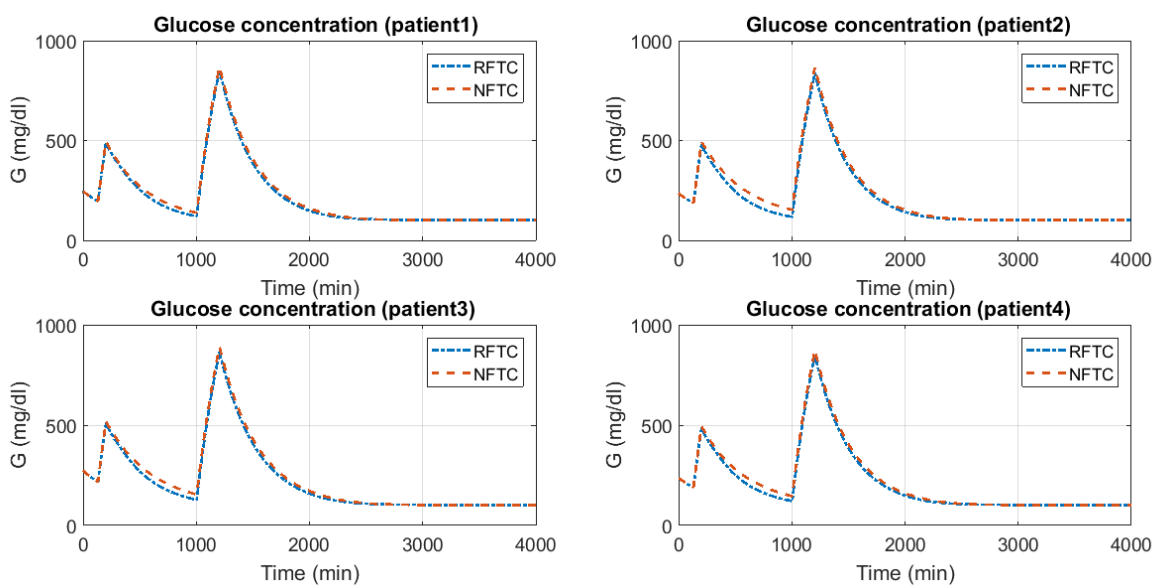
روش استفاده شده	مجموع دوز داروی تزریقی (mg)
یادگیری تقویتی	3.1432×10^4
شبکه‌های عصبی	5.8819×10^4

با توجه به جدول (۶-۲)، مجموع دوز داروی تزریق شده با استفاده از روش یادگیری تقویتی نسبت به روش شبکه‌های عصبی بسیار کمتر می‌باشد.

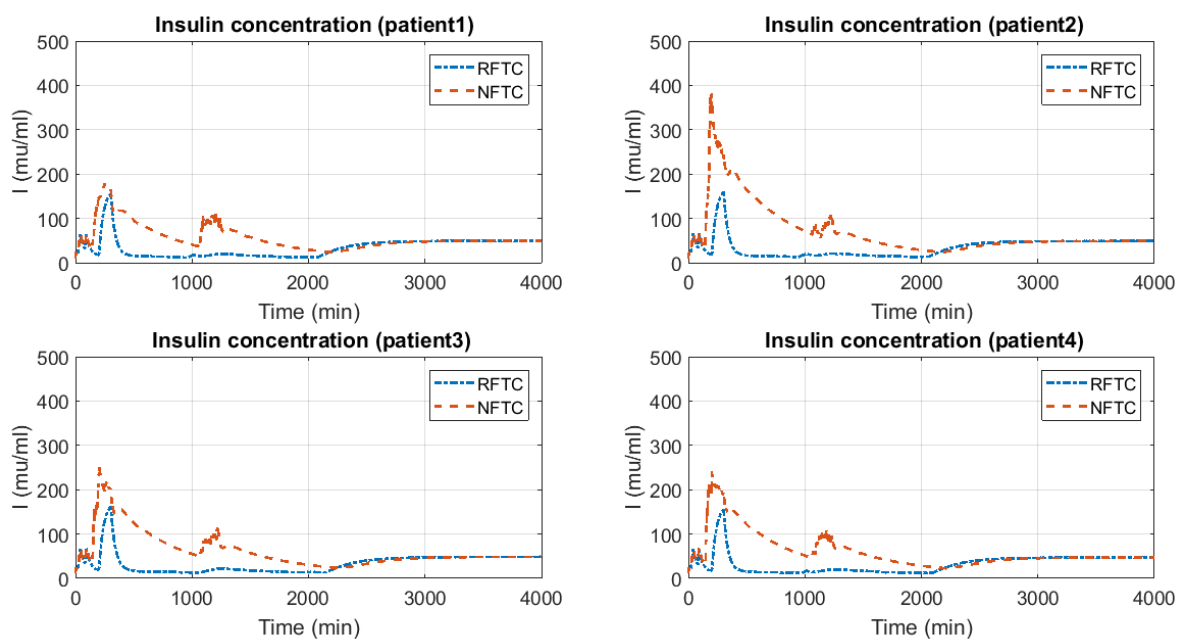
۶-۴-۳- اعمال عدم قطعیت در پارامترهای سیستم و شرایط اولیه

در این بخش کنترل میزان گلوکز خون توسط هر دو روش ذکر شده برای مدل ریاضی چهار بیمار دیگر مبتلا به دیابت بررسی می‌شود. برای این کار ابتدا، با لحاظ کردن عدم قطعیت در پارامترهای سیستم و شرایط اولیه، مقدار آن‌ها برای چهار بیمار دیگر محاسبه خواهد شد. این محاسبه به صورت در نظر گرفتن پنج درصد تغییرات در شرایط اولیه و ضرایب ثابت بیمار اصلی خواهد بود. تغییرات میزان گلوکز برای هر چهار بیمار در حضور وقوع عیب در سیستم، با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی در شکل (۶-۱۲) قابل مشاهده می‌باشد. شکل (۶-۱۳) میزان انسولین خون را در این چهار بیمار نشان می‌دهد. شکل (۶-۱۴) میزان دوز داروی تزریقی به چهار بیمار را با استفاده از هر دو روش انتخابی نشان می‌دهد. همانطور که در شکل (۶-۱۲) مشاهده می‌شود، روش مسیرهای شایستگی با سرعت بیشتر توانسته است میزان گلوکز خون را در دیگر بیماران مبتلا به دیابت کنترل کرده و به مقدار مطلوب برساند.

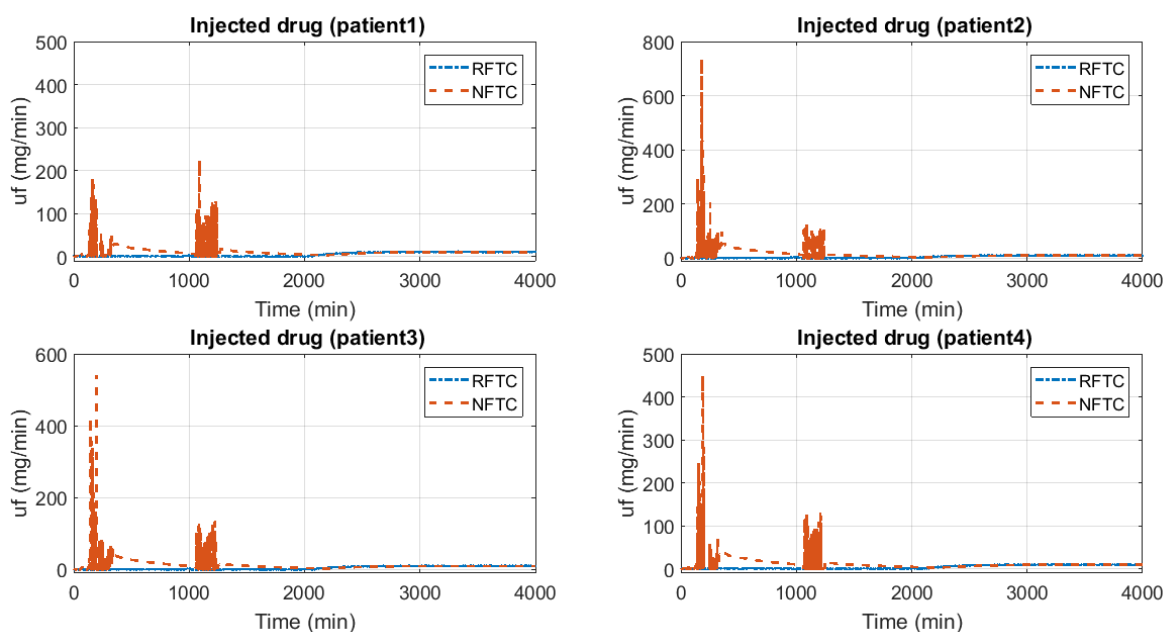
با توجه به شکل (۶-۱۳)، میزان انسولین در چهار بیمار با استفاده از روش مسیرهای شایستگی کمتر از حالتی می‌باشد که از شبکه‌های عصبی توابع پایه شعاعی استفاده شده است. با وجود این که در بعضی از مواقع میزان انسولین کمتر از مقدار مطلوب می‌باشد، اما باز هم هر دو روش انتخابی قادر به کنترل آن و رساندن آن به مقدار مطلوب بوده‌اند. همانطور که در شکل (۶-۱۴) نشان داده شده است، میزان دوز داروی تزریقی به هر چهار بیمار با استفاده از روش مسیرهای شایستگی بسیار کمتر از دوز داروی تزریقی توسط روش شبکه‌های عصبی توابع پایه شعاعی می‌باشد.



شکل ۶-۱۱: میزان گلوکز خون در چهار بیمار دیابتی مختلف با استفاده از مسیرهای شایستگی و شبکه‌های عصبی



شکل ۶-۱۲: میزان انسولین خون در چهار بیمار دیابتی مختلف با استفاده از مسیرهای شایستگی و شبکه‌های عصبی



شکل ۶-۱۳: میزان میزان دوز داروی تزریقی به چهار بیمار دیابتی مختلف با استفاده از مسیرهای شایستگی و شبکه‌های عصبی

۶-۴-۴- اعمال نويز به سيستم در حضور عدم قطعيت در پارامترهاي سيستم و شرايط اوليه

به دليل اينکه شرايط محيطی ثابت نيست و هر لحظه در حال تغيير می‌باشد، در اين بخش با اعمال نويز به سيستم رفتار بررسی خواهد شد. در اين حالت معادلات سيستم با يك مقدار تصادفی که نويز سفید می‌باشد و با متغير نمايش داده شده است، جمع می‌شوند و روابط جديد به صورت معادلات (۴۸-۶) تا (۵۰-۶) خواهد بود.

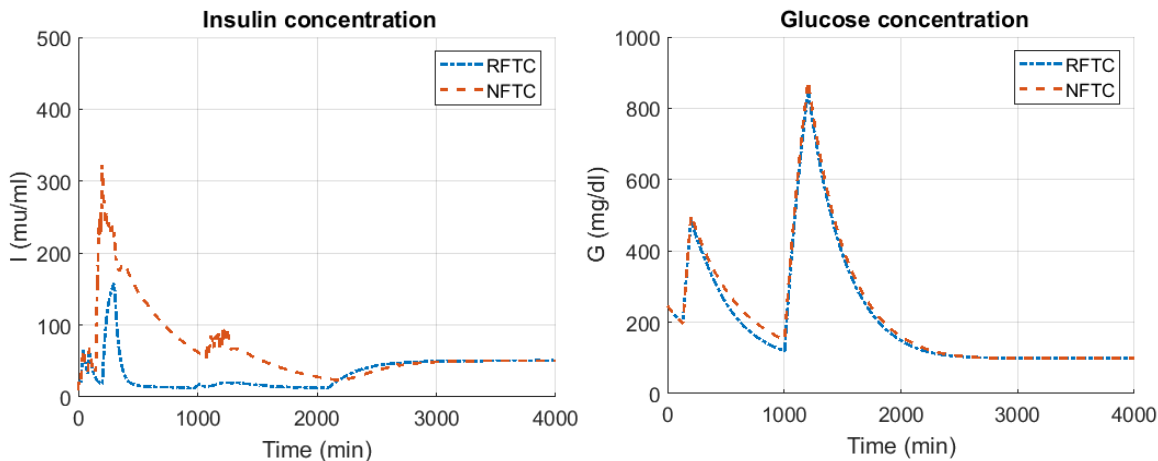
$$\frac{dG}{dt} = -m_1 G + m_2 I + m_1 G_b + w_1 \quad (۴۸-۶)$$

$$\frac{dC}{dt} = -m_2 X + m_3 I - m_3 I_b + m_6 I_b \quad (۴۹-۶)$$

$$\frac{dI}{dt} = -m_3 I + m_4 G + m_4 m_5 - m_6 I + m_6 I_b + w_2 \quad (۵۰-۶)$$

طبق روابط فوق، نويز وارد شده به سيستم فقط به متغيرهاي حالت در يادگيري تقويتي اعمال می‌شود. میزان گلوکز در حضور وقوع عيب در سيستم و نويز وارد شده، برای بیمار اول در شکل (۱۵-)

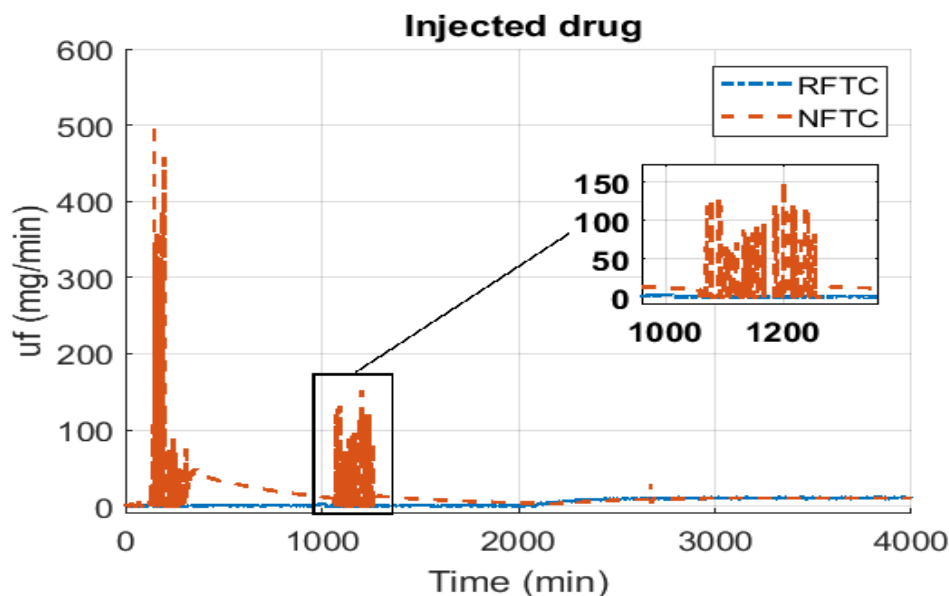
۶) قابل نمایش می‌باشد. میزان تغییرات انسولین در بیمار اول نیز در شکل (۶-۱۶) نمایش داده شده است.



شکل ۶-۱۵: انسولین خون در حضور عیب و نویز در بیمار مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی

شکل ۶-۱۴: گلوکز خون در حضور عیب و نویز در بیمار مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی

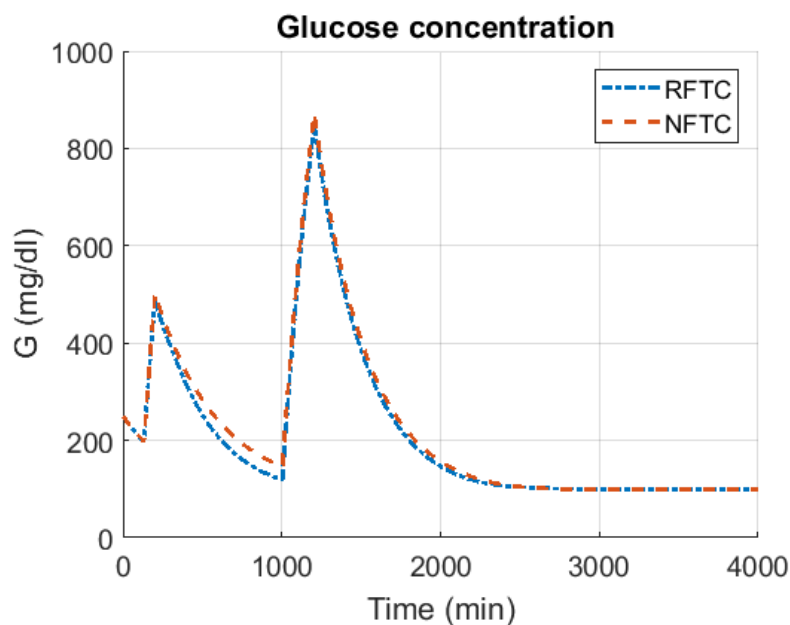
با توجه به شکل (۶-۱۵)، هر دو روش انتخابی قادر به کنترل میزان گلوکز و رساندن به مقدار مطلوب در حضور نویز اعمالی به سیستم، بوده‌اند. اما رفتار کنترلر طراحی شده با مسیرهای شایستگی کمی بهتر می‌باشد. همان‌طور که در شکل (۶-۱۶) نشان داده شده است، در حضور نویز و با استفاده از روش مسیرهای شایستگی میزان انسولین تزریق شده به بیمار کمتر از حالتی می‌باشد که از روش شبکه‌های عصبی توابع پایه شعاعی استفاده شده است و با وجود اینکه در هر دو روش انتخابی میزان انسولین کمتر از مقدار مطلوب شده است، اما باز هم توانسته‌اند آن را کنترل کرده و به مقدار مطلوب برسانند. شکل (۶-۱۷) میزان دوز داروی تزریق شده به بیمار را نشان می‌دهد.



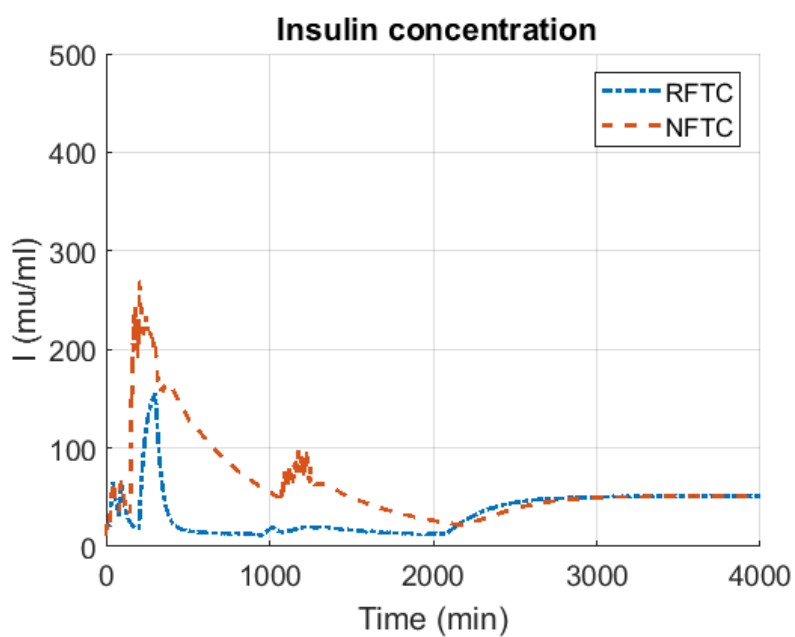
شکل ۶-۱۶: دوز داروی تزریقی به بیمار مبتلا به دیابت در حضور عیب و نویز با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی
با توجه به شکل (۶-۱۷)، میزان دوز داروی تزریقی توسط روش مسیرهای شایستگی بسیار کمتر از حالت استفاده از روش شبکه‌های عصبی توابع پایه شعاعی می‌باشد.

۶-۴-۵- اعمال عدم قطعیت در پارامترهای سیستم در طول درمان

پارامترهای بدن یک بیمار به دلیل فعالیت‌های روزمره وی مرتب در حال تغییر می‌باشد و ثابت نیست. در این بخش برای سیستم اصلی (بیمار اصلی) تغییرات پارامترها در لحظه لحاظ شده است و مقدار آن‌ها در هر لحظه به طور تصادفی دو درصد از مقدار اصلی کم و یا زیاد می‌شوند. شکل (۶-۱۸) میزان گلوکز برای بیمار مبتلا به دیابت، در حضور عدم قطعیت در پارامترهای سیستم را نشان می‌دهد. همانطور که در شکل (۶-۱۸) نشان داده شده است، با وجود اینکه پارامترهای بدن بیمار در هر لحظه تغییر می‌کند و ثابت نیست، اما هر دو روش انتخابی باز هم قادر به کنترل میزان گلوکز و نگه داشتن آن روی مقدار مطلوب در حضور عیب وارد شده به سیستم هستند. تغییرات میزان انسولین نیز در شکل (۶-۱۹) قابل نمایش می‌باشد.

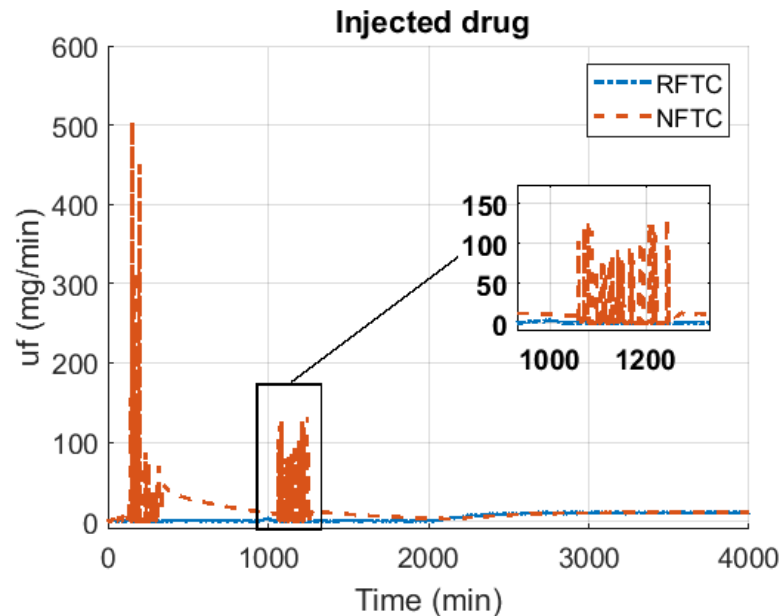


شکل ۶-۱۷: گلوکز خون در حضور عدم قطعیت در پارامترهای سیستم در طول درمان در رخدادهای عیب در بیمار مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی



شکل ۶-۱۸: انسولین خون در حضور عدم قطعیت در پارامترهای سیستم در طول درمان در رخدادهای عیب در بیمار مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی
با توجه به شکل (۶-۱۹)، هر دو روش انتخابی قادر به کنترل میزان انسولین در حضور عدم قطعیت در پارامترهای سیستم در هر لحظه، می‌باشند. با وجود اینکه میزان انسولین کمتر از مقدار مطلوب شده

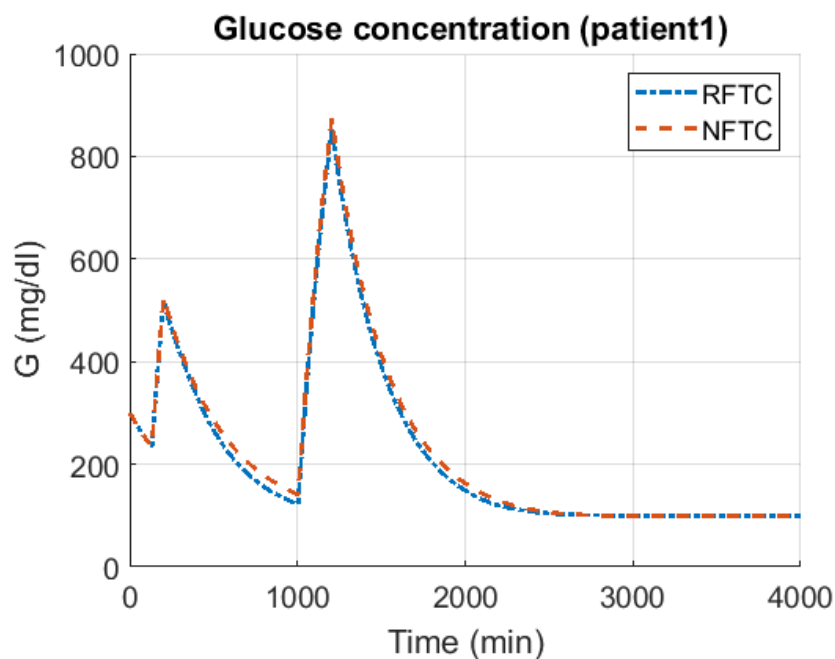
است، اما باز هم هر دو روش توانسته‌اند آن را کنترل کرده و به مقدار مطلوب برسانند. شکل (۶-۲۰) میزان دوز داروی تزریقی به بیمار را در این حالت نشان می‌دهد.



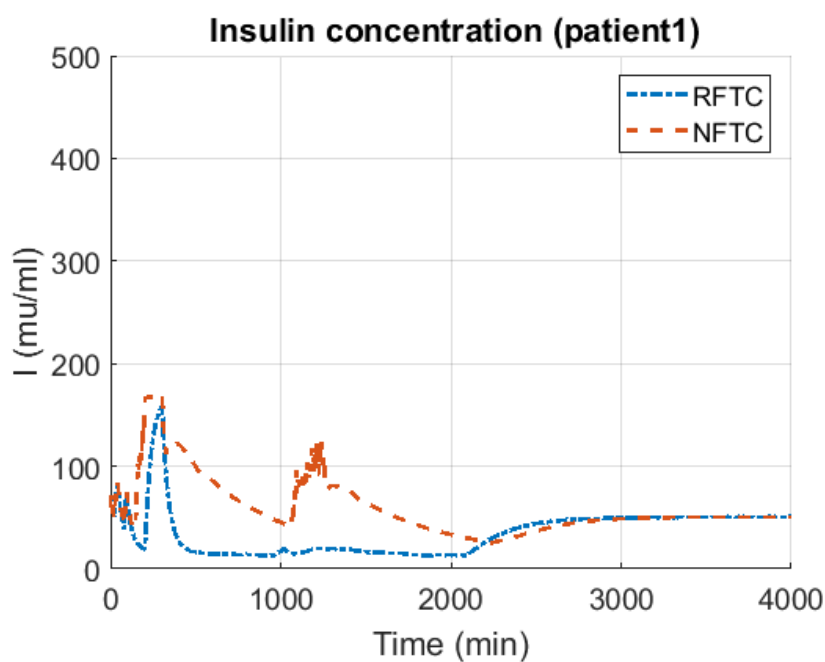
شکل ۶-۱۹: میزان دوز داروی تزریقی در حضور عدم قطعیت در پارامترهای سیستم در طول درمان در رخداد عیب در بیمار مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی با توجه به شکل (۶-۲۰)، میزان دوز داروی تزریقی توسط روش مسیرهای شایستگی بسیار کمتر از حالت استفاده از روش شبکه‌های عصبی توابع پایه شعاعی می‌باشد.

۶-۴-۶ اعمال تغییرات در شرایط اولیه بیمار

در این بخش برای بیمار شماره ۱ تغییرات چشم‌گیری روی شرایط اولیه انجام گرفت که مقادیر آن‌ها به ترتیب $G_0 = 300, X_0 = 60, I_0 = 75$ می‌باشد و کنترل میزان گلوکز خون با استفاده از هر دو روش بیان شده بررسی می‌شود. شکل (۶-۲۱) تغییرات میزان گلوکز خون را در این حالت نشان می‌دهد. همانطور که در شکل (۶-۲۱) نشان داده شده است، با وجود تغییرات چشم‌گیر در شرایط اولیه بیمار، باز هم هر دو روش انتخابی قادر به کنترل میزان گلوکز و رساندن آن به مقدار مطلوب بوده‌اند. شکل (۶-۲۲) تغییرات میزان انسولین را در حالت تغییرات در شرایط اولیه نشان می‌دهد.



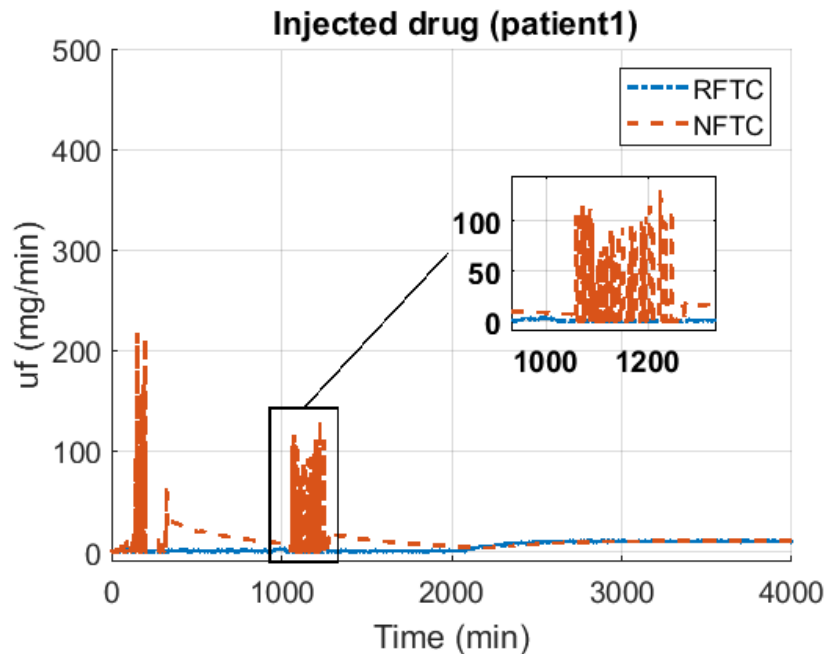
شکل ۶-۲۰: گلوکز خون در حضور تغییر شرایط اولیه و رخداد عیب در بیمار شماره ۱ مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی



شکل ۶-۲۱: انسولین خون در حضور تغییر شرایط اولیه و رخداد عیب در بیمار شماره ۱ مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی

با توجه به شکل (۶-۲۲) روش مسیرهای شایستگی با وجود تغییر در شرایط اولیه سیستم، باز هم با سرعت بیشتر و تزریق انسولین کمتر توانسته است میزان گلوکز را کنترل کرده و به مقدار مطلوب

برساند. میزان دوز داروی تزریقی به بیمار مبتلا به دیابت در این حالت، در شکل (۲۳-۶) قابل مشاهده می‌باشد.



شکل ۶-۲۲: میزان دوز داروی تزریقی در حضور تغییر شرایط اولیه و رخداد عیب در بیمار شماره ۱ مبتلا به دیابت با استفاده از روش مسیرهای شایستگی و شبکه‌های عصبی توابع پایه شعاعی
با توجه به شکل فوق، میزان دوز داروی تزریقی توسط روش مسیرهای شایستگی بسیار کمتر از حالت استفاده از روش شبکه‌های عصبی توابع پایه شعاعی می‌باشد.

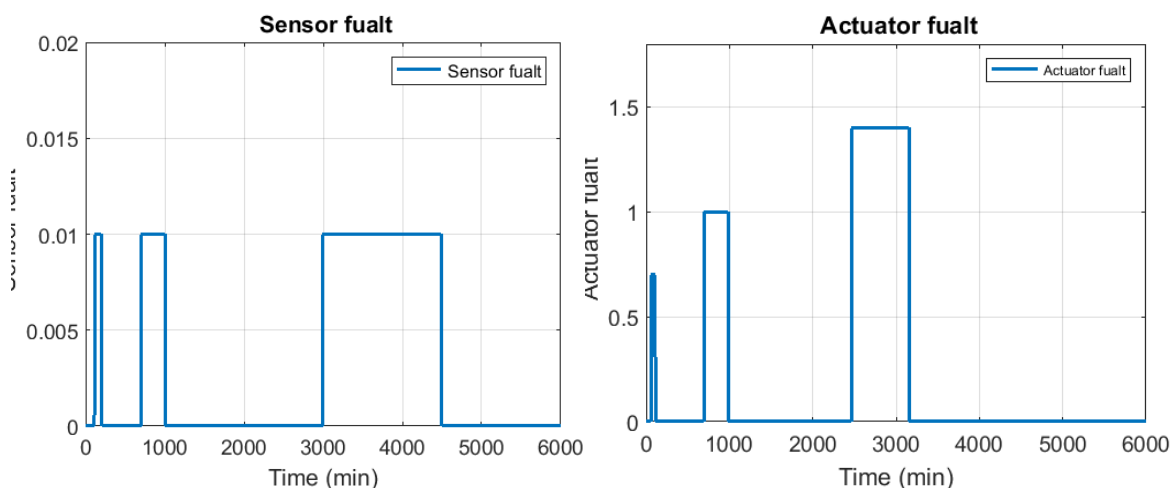
۶-۵- شبیه‌سازی ساختارهای کنترلی فوق بر روی مدل تاخیر دار پالمبو

اکنون روش‌های مذکور در این فصل بر روی مدل تاخیر دار پالمبو آورده شده است. روابط این مدل در فصل ۴ بیان گردیده است. ثوابتی که برای سیستم در نظر گرفته شده است به صورت جدول ۶-۳ می‌باشد. شرایط اولیه که برای بیمار در نظر گرفته شده است، بدین ترتیب می‌باشد: $G_0 = 98.6056, I_0 = 6.1915$

جدول ۶-۳: ثوابت سیستم پالمبو [۵۹]

Parameter	Value	Parameter	Value
G_b	6.14	V_g	0.187
I_b	93.669	K_{xi}	1.211×10^{-2}
T_{igmax}	1.573	T_{gh}	0.003
δ	3.205	V_i	0.25
G^*	9	K_{xgi}	3.11×10^{-5}
τ_g	24		

مقادیر مطلوب برای متغیرها که در نهایت می‌بایست به آن برسند، برای گلوکز 5.6 mM و انسولین 96PM می‌باشد. دوز داروی تزریقی که متغیر u^f می‌باشد، باید به گونه‌ای تعیین شود که هنگام بروز عیب در سیستم که در سنسور و عملگر اتفاق می‌افتد، روش انتخابی قادر به کنترل میزان گلوکز و نگه‌داشتن آن در سطح مطلوب باشد. به دلیل وجود تاخیر در سیستم و وابستگی متغیرها در هر لحظه به چندین لحظه قبل، حل معادلات دیفرانسیل با استفاده از دستور 'dde' انجام می‌شود. این دستور برای حل معادلات دیفرانسیل دارای تاخیر در متلب می‌باشد. عیوب وارد شده به عملگر و سنسور سیستم به ترتیب در شکل‌های (۶-۲۴) و (۶-۲۵) نمایش داده شده است.



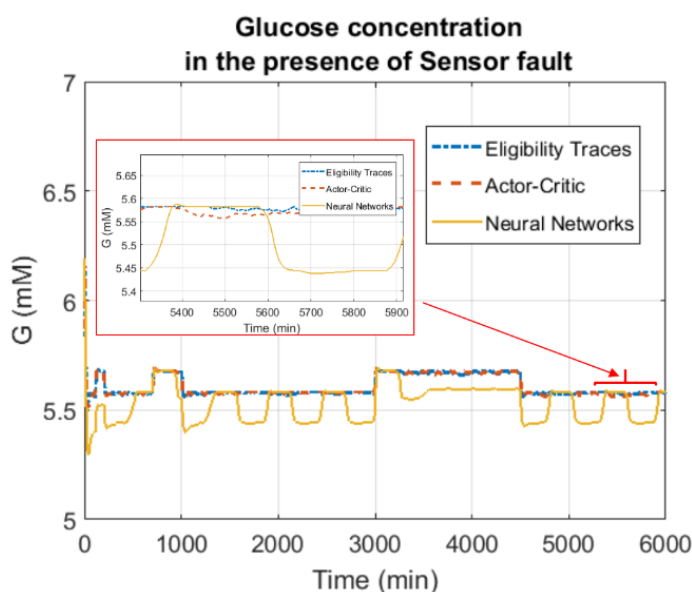
شکل ۶-۲۴: عیب وارد شده به سنسور سیستم

شکل ۶-۲۳: عیب وارد شده به عملگر سیستم

¹ delay differential equation

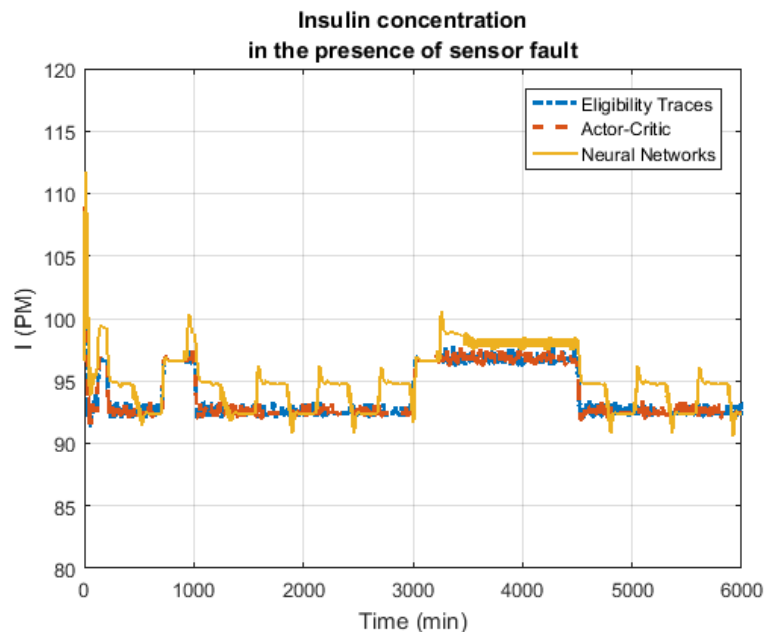
۶-۵-۱- عیب وارد شده به سنسور

در این بخش، عیب وارد شده به سیستم تنها به سنسور آن در نظر گرفته شده است. شکل (۶-۲۶) کنترل میزان گلوکز را در یک بیمار مبتلا به دیابت در حضور عیب سنسور نمایش می‌دهد. از روش مسیرهای شایستگی، بازیگر-نقاد و شبکه‌های عصبی [۳۴] برای کنترل میزان گلوکز استفاده شده است.



شکل ۶-۲۵: میزان گلوکز در یک بیمار مبتلا به دیابت در حضور عیب سنسور با استفاده از روش‌های مسیرهای شایستگی، بازیگر-نقاد و شبکه‌های عصبی

همان‌طور که در شکل فوق مشاهده می‌شود، روش مسیرهای شایستگی با تزریق دوز داروی کمتر و با وجود رخداد عیب در سنسور سیستم، توانسته است مقدار گلوکز را به مقدار مطلوب برساند و روی این مقدار نگه دارد. روش بازیگر-نقاد هم تقریباً عملکردی مشابه با روش مسیرهای شایستگی دارد اما روش مبتنی بر شبکه‌های عصبی نتوانسته است میزان گلوکز را به مقدار مطلوب برساند، و این روش قادر به کنترل سریع آن نبوده است. مجموع دوز داروی تزریقی در دو روش شبکه‌های عصبی و بازیگر-نقاد بسیار بیشتر از روش مسیرهای شایستگی می‌باشد که این امر اثرات زیان‌بار دارو را به همراه خواهد داشت. شکل (۶-۲۷) نمایش‌گر میزان انسولین برای کنترل میزان گلوکز با استفاده از روش مسیرهای شایستگی، بازیگر-نقاد و شبکه‌های عصبی می‌باشد.

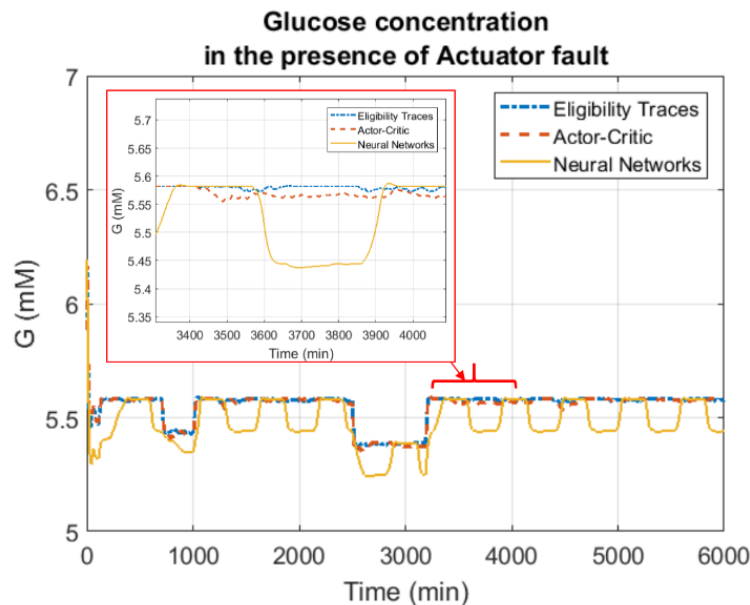


شکل ۶-۲۶: میزان انسولین در یک بیمار مبتلا به دیابت در حضور عیب سنسور با استفاده از روش‌های مسیرهای شایستگی، بازیگر-نقاد و شبکه‌های عصبی

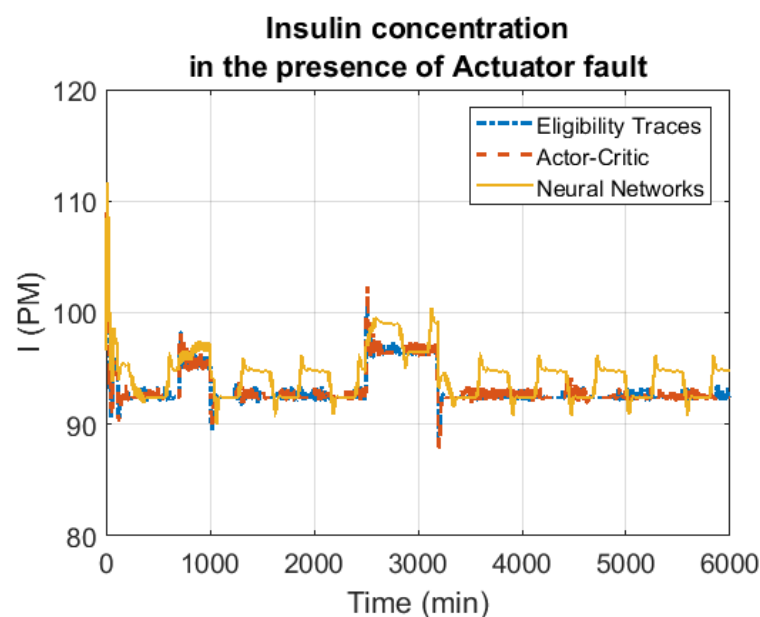
با توجه به شکل فوق، میزان انسولین در هنگام رخداد عیب با استفاده از روش مسیرهای شایستگی بسیار سریع کنترل شده و روی مقدار مطلوب نگه داشته شده است. در روش بازیگر-نقاد هم تقریباً رفتار مشابه‌ای از خود نشان داده است اما در روش تزریق دوز با استفاده از شبکه‌های عصبی در حضور عیب رفتار مناسبی از خود نشان نداده است.

۶-۵-۲- عیب وارد شده به عملگر

در این بخش، عیب وارد شده به سیستم تنها به عملگر آن در نظر گرفته شده است. شکل (۶-۲۸) کنترل میزان گلوکز را در یک بیمار مبتلا به دیابت در حضور عیب وارد شده به عملگر را نمایش می‌دهد. از روش مسیرهای شایستگی، بازیگر-نقاد و شبکه‌های عصبی برای کنترل میزان گلوکز استفاده شده است.



شکل ۶-۲۷: میزان گلوکز در یک بیمار مبتلا به دیابت در حضور عیب وارد شده به عملگر با استفاده از روش‌های مسیره‌های شایستگی، بازیگر- نقاد و شبکه‌های عصبی

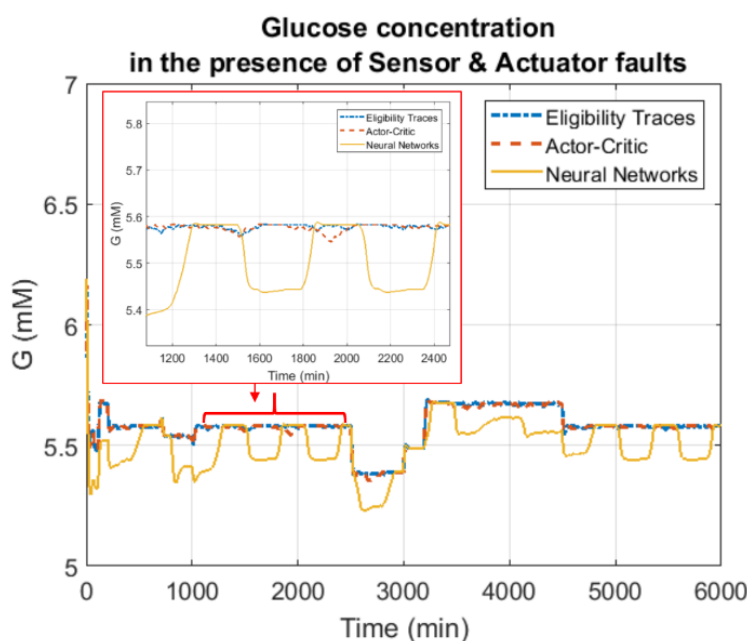


شکل ۶-۲۸: میزان انسولین در یک بیمار مبتلا به دیابت در حضور عیب وارد شده به عملگر سیستم با استفاده از روش‌های مسیره‌های شایستگی، بازیگر- نقاد و شبکه‌های عصبی

همانطور که در شکل (۶-۲۸) نشان داده شده است، روش مسیره‌های شایستگی در هنگام رخداد عیب در عملگر سیستم نیز قادر به کنترل میزان گلوکز و رساندن آن به مقدار مطلوب بوده است. در حالی که در روش تزریق دوز داروی ثابت میزان میزان گلوکز کنترل نشده و کمتر مقدار مطلوب آن می‌باشد. شکل (۶-۲۹) میزان انسولین خون را هنگام رخداد عیب در عملگر سیستم نشان می‌دهد.

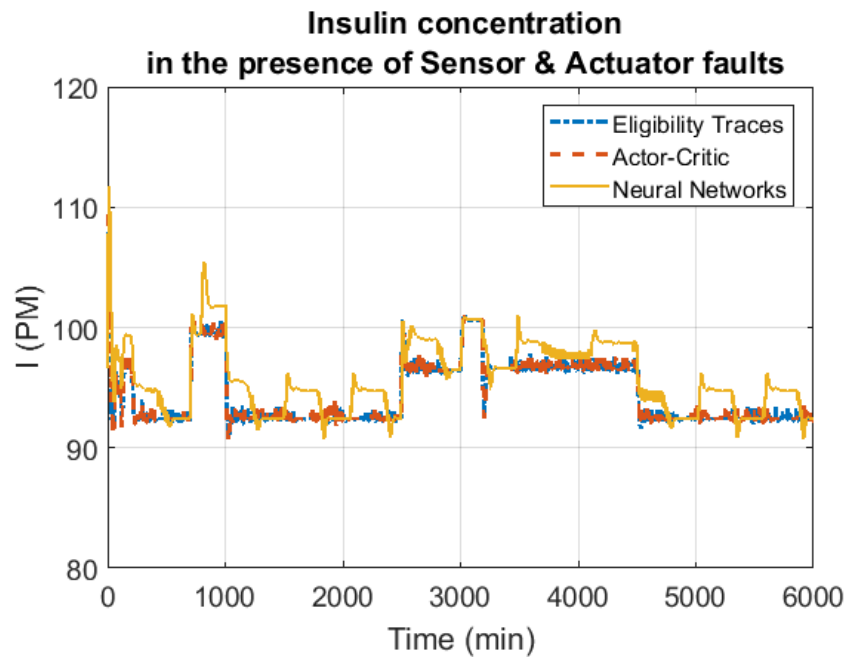
۳-۵-۶- عیب وارد شده به سنسور و عملگر

در این بخش، عیب وارد شده به سنسور و عملگر به صورت همزمان در نظر گرفته شده است. شکل (۳۰-۶) کنترل میزان گلوکز را در یک بیمار مبتلا به دیابت در حضور عیب وارد شده به سنسور و عملگر را نمایش می‌دهد. ز روش مسیرهای شایستگی، بازیگر-نقاد و شبکه های عصبی برای کنترل میزان گلوکز استفاده شده است.



شکل ۳۰-۶: میزان گلوکز در یک بیمار مبتلا به دیابت در حضور عیب سنسور و عملگر با استفاده از روشهای مسیرهای شایستگی، بازیگر-نقاد و شبکه‌های عصبی

همانطور که در شکل فوق نشان داده شده است، در هنگام رخداد هر دو عیب در سیستم، روش مبتنی بر شبکه‌های عصبی قادر به کنترل میزان گلوکز خون و رساندن آن به مقدار مطلوب نبوده است و همچنین در بازه رخداد عیب نیز سرعت این روش در کنترل گلوکز بسیار پایین بوده است. اما روش مسیرهای شایستگی میزان گلوکز را با سرعت بیشتر و تزریق دوز داروی کمتر کنترل کرده است و روی مقدار مطلوب نگه داشته است. شکل (۳۱-۶) میزان انسولین خون را توسط استفاده از هر سه روش ذکر شده نشان می‌دهد.



شکل ۶-۳۰: میزان انسولین در یک بیمار مبتلا به دیابت در حضور عیب وارد شده به سنسور و عملگر سیستم با استفاده از روش‌های مسیرهای شایستگی، بازیگر-نقاد و شبکه‌های عصبی

همانطور که در شکل فوق نشان داده شده است، در حضور عیب در سنسور و عملگر سیستم، روش مسیرهای شایستگی میزان انسولین را کنترل کرده و روی مقدار مطلوب نگه داشته است. این روش برخلاف روش مبتنی بر شبکه‌های عصبی از افزایش خیلی زیاد انسولین در بازه رخداد عیب نیز جلوگیری کرده است.

مجموع دوز داروی تزریقی در حالت بروز هر کدام از عیوب وارد شده به سیستم با استفاده از روش مسیرهای شایستگی، تزریق دوز داروی با استفاده از روش بازیگر-نقاد و تزریق دوز دارو با استفاده از شبکه‌های عصبی در تمام مدت کنترل میزان گلوکز برای بیمار مبتلا به دیابت در جدول (۶-۴) بیان شده است.

جدول ۴-۶: مجموع دوز داروی تزریقی با استفاده از روش یادگیری تقویتی و تزریق دوز داروی ثابت

سیگنال کنترلی (دوز دارو)	روش	عیب وارد شده به سیستم	مجموع دوز داروی تزریقی
u^f	مسیرهای شایستگی	سنسور	19.01 mg
		عملگر	18 mg
		سنسور و عملگر	18.01 mg
	بازیگر-نقاد	سنسور	20.48mg
		عملگر	34.05 mg
		سنسور و عملگر	25.64mg
	شبکه‌های عصبی	سنسور	357.8235mg
		عملگر	325.1007mg
		سنسور و عملگر	368.1947mg

با توجه به جدول (۴-۶)، مجموع دوز داروی تزریق شده با استفاده از روش یادگیری تقویتی (مسیرهای شایستگی) در حالت رخداد هر سه عیب در سیستم نسبت به نسبت به تزریق دوز دارو با استفاده از روش یادگیری تقویتی (بازیگر-نقاد) کمتر می‌باشد و این دو روش از روش شبکه‌های عصبی دوز بسیار کمتری برای درمان استفاده کرده‌اند.

فصل ۷- کنترل تحمل‌پذیر عیب با استفاده از روش مسیرهای

شایستگی و در نظر گرفتن پنجره زمانی

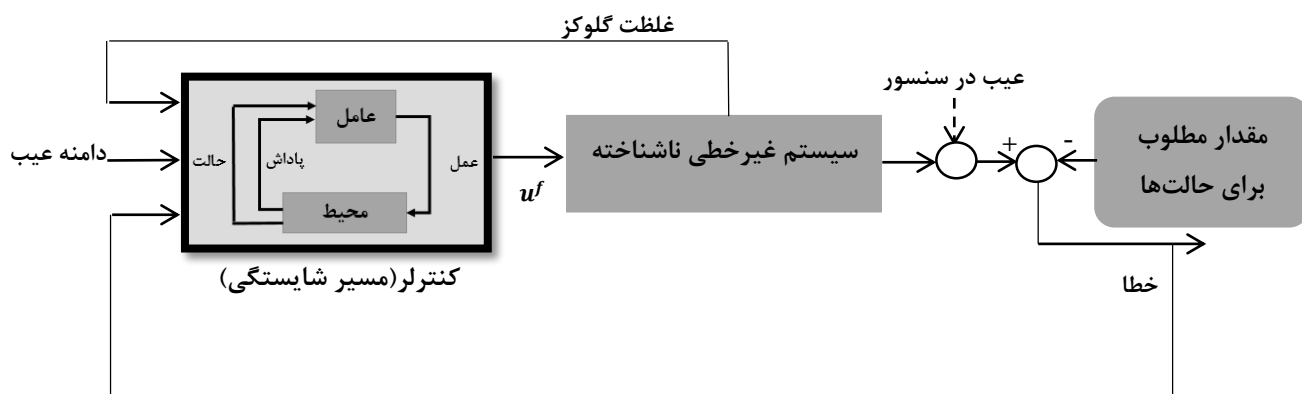
۷-۱- مقدمه

در روش‌هایی که در فصل پنج و شش مطرح شد کنترل تحمل‌پذیر عیب برای حالتی انجام می‌شد که عیب وارد شده به سنسور و یا عملگر به صورت ناگهانی باشد. در این فصل با استفاده از مفهومی تحت عنوان پنجره زمانی قادر خواهیم بود عیب‌های نرم را نیز تشخیص دهیم. علاوه بر این مورد، در این فصل تنها از یادگیری تقویتی به عنوان کنترل‌کننده تحمل‌پذیر عیب استفاده شده است و کنترل‌کننده مد لغزشی که در فصل شش توضیح داده شد حذف گردیده است. در انتها نشان داده خواهد شد که ساختار پیشنهادی این فصل با وجود سادگی ساختاری، رفتار بسیار بهتری در کنترل تحمل‌پذیر عیب دارا می‌باشد.

۷-۲- کنترل تحمل‌پذیر عیب با استفاده از روش مسیرهای شایستگی و در نظر گرفتن

پنجره زمانی

یکی از روش‌های دیگری که در این رساله برای کنترل تحمل‌پذیر عیب در نظر گرفته شده است به صورت شکل (۷-۱) می‌باشد. حالت‌های در نظر گرفته شده برای مسیرهای شایستگی شامل مقدار متغیرهای سیستم، خطا (اختلاف میان مقدار مطلوب و مقدار بدست آمده از سیستم در حالتی که سیستم دچار عیب شده باشد) و میزان دامنه عیب وارد شده به سیستم در هر گام زمانی، می‌باشد. میزان μ^f به عنوان عمل برای عامل در نظر گرفته شده است. شکل (۶-۶) کنترل گلوکز با استفاده از روش مسیرهای شایستگی را نشان می‌دهد.



شکل ۱-۷: کنترل میزان گلوکز با استفاده از روش مسیرهای شایستگی در زمان رخداد عیب در سنسور
 انتخاب اعمال به صورت احتمالاتی بوده و توسط روش ϵ -حریصانه انجام گرفته است. می باشد که بر اساس رابطه (۱-۷) محاسبه می شود.

$$a_t = \begin{cases} a_t^* & \text{with probability } 1 - \epsilon \\ \text{random action} & \text{with probability } \epsilon \end{cases} \quad (1-7)$$

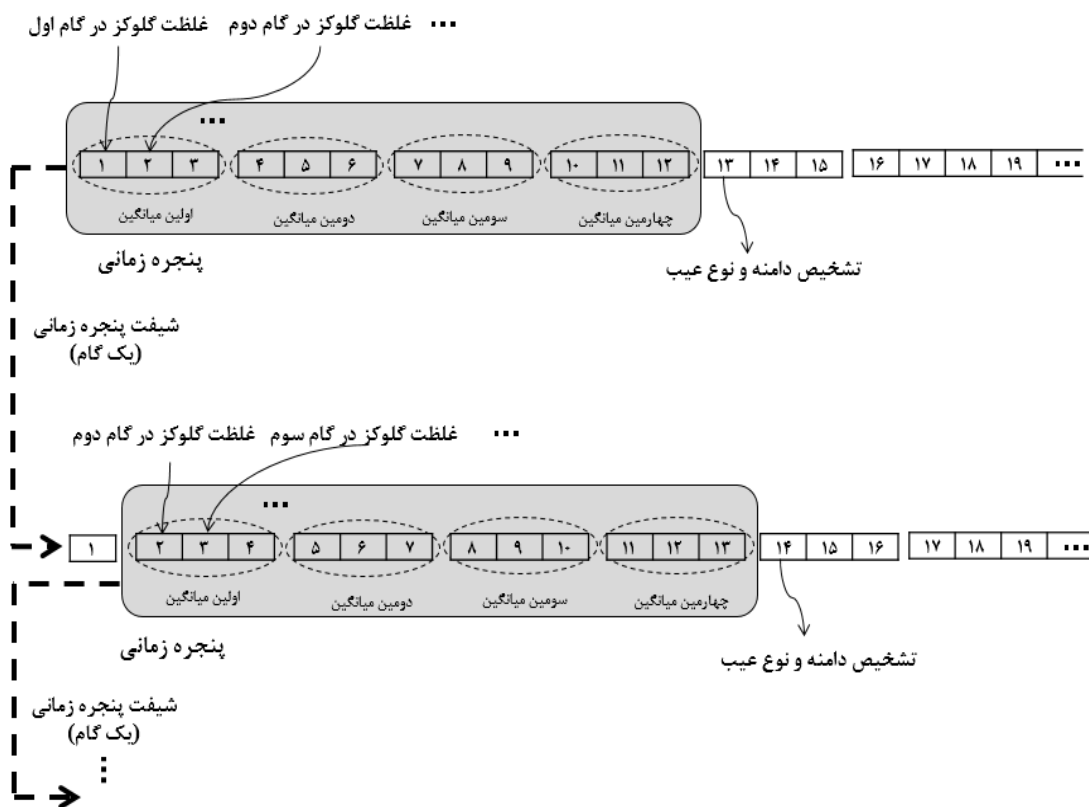
طبق رابطه فوق، بهترین عمل (a_t^*) با احتمال $1 - \epsilon$ و بقیه اعمال با احتمال یکسان انتخاب می شوند. در شکل (۲-۷) روش انتخاب اعمال توسط این سیاست قابل نمایش می باشد.

Action	a_1	a_2	a_3^*	a_4
Action Value	0.01	0.2	1.3	0.04
Selection Probability	$\frac{\epsilon}{n-1}$	$\frac{\epsilon}{n-1}$	$1 - \epsilon$	$\frac{\epsilon}{n-1}$

شکل ۲-۷: انتخاب اعمال توسط سیاست ϵ -greedy
 با توجه به شکل (۲-۷)، عمل a_3 بهترین عمل با بالاترین ارزش بوده است، پس با احتمال $1 - \epsilon$ انتخاب می شود و بقیه اعمال با احتمال یکسان و مقدار $\frac{\epsilon}{n-1}$ انتخاب می شوند که n تعداد اعمال قابل انجام می باشد.

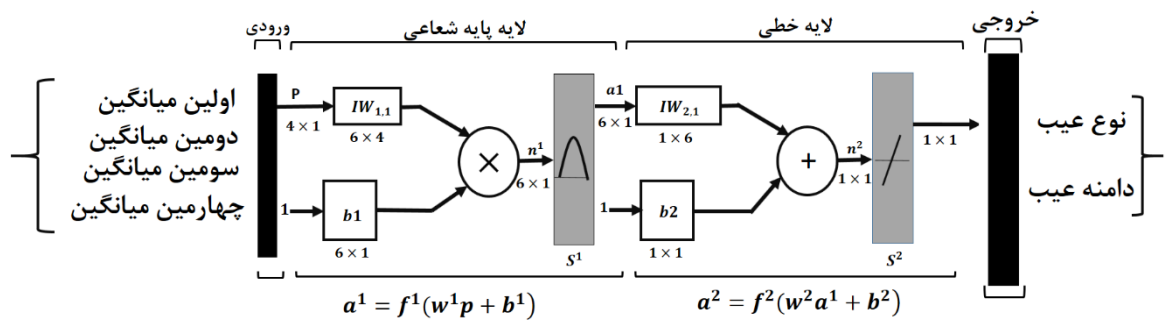
عیوب وارد شده به سنسور سیستم دو نوع بوده و از نوع نرم و ناگهانی می باشد.
 برای تشخیص دامنه و نوع عیب وارد شده به سیستم از روش شبکه های عصبی پیشرو و پنجره زمانی استفاده شده است. برای تشخیص دامنه و نوع عیب وارد شده، بدین صورت عمل شده است.

دوازده مقدار متوالی حالت‌های سیستم در پنجره زمانی قرار می‌گیرند. بین هر سه مقدار میانگین گرفته می‌شود و چهار مقدار متوالی از میانگین‌گیری‌ها بدست می‌آید. این چهار مقدار بر اساس نوع عیب وارد شده و دامنه آن برچسب زده می‌شوند. در لحظه بعد پنجره زمانی یک گام به سمت جلو شیفت پیدا می‌کند و موارد فوق مجدد تکرار می‌گردد. به همین ترتیب داده‌های آموزشی برای آموزش شبکه عصبی بدست می‌آید که برای یادگیری به شبکه عصبی داده می‌شوند. بر اساس یادگیری انجام شده توسط روش شبکه‌های عصبی، در حالت کنترل میزان گلوکز با استفاده از روش یادگیری تقویتی تشخیص دامنه و نوع عیب انجام می‌شود. روند تشخیص عیب و حرکت پنجره زمانی در شکل (۷-۳) قابل نمایش می‌باشد.



شکل ۷-۳. پنجره زمانی در نظر گرفته شده برای تشخیص عیب

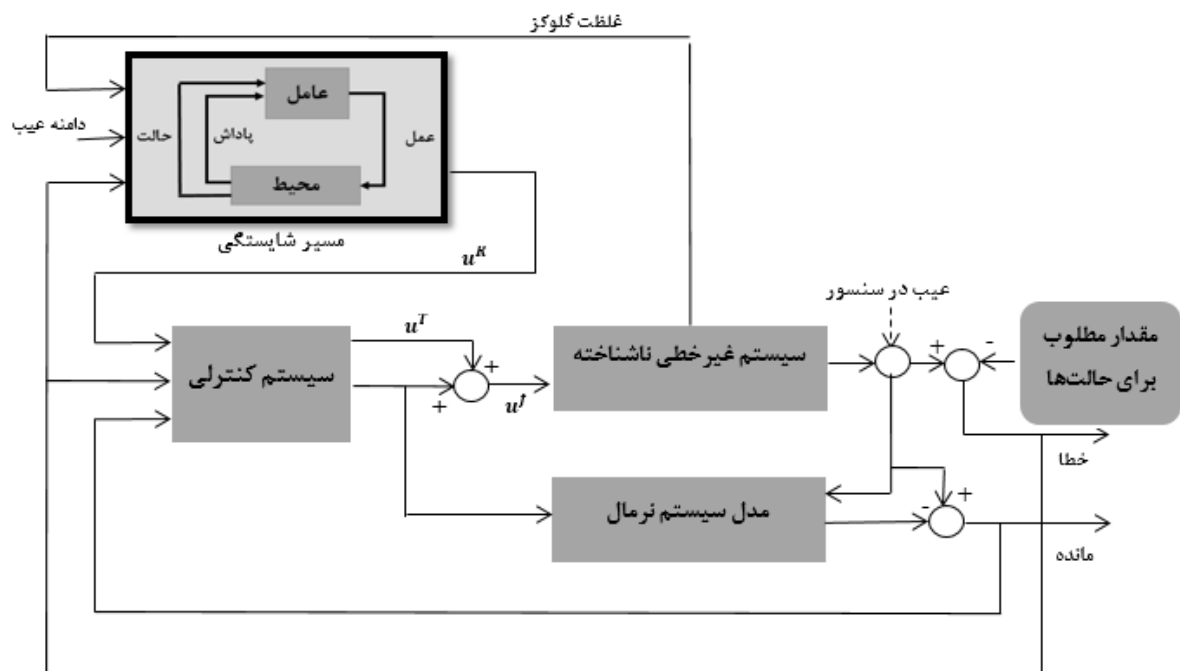
لازم به ذکر می‌باشد که میزان دامنه عیب وارد شده به عنوان یکی از حالت‌ها در روش یادگیری تقویتی در نظر گرفته می‌شود که بر روی میزان دوز داروی تزریقی به بیمار مبتلا به دیابت تاثیرگذار خواهد بود. نمای کلی شبکه عصبی مورد استفاده در شکل (۷-۴) نشان داده شده است



شکل ۴-۷: نمای کلی شبکه عصبی جهت آموزش برای تشخیص دامنه و نوع عیب

۷-۳- شبیه سازی

در این قسمت روش پیشنهادی در این فصل با روشی مشابه فصل ۶ مقایسه کرده ایم با این تفاوت که دامنه عیب شناسایی شده به کمک شبکه عصبی را نیز به عنوان یکی از ورودی‌ها برای قسمت مسیر شایستگی در نظر گرفته‌ایم شمای کلی کار در شکل (۷-۵) نشان داده شده است.



شکل ۷-۵: کنترل میزان گلوکز با استفاده از ترکیب مسیرهای شایستگی و کنترل مد لغزشی در زمان رخداد عیب در سنسور (حالت بهبود یافته فصل ۶)

در شکل فوق، u^R عمل انتخاب شده توسط روش مسیرهای شایستگی در مسئله یادگیری تقویتی می‌باشد که انتخاب صحیح و مقدار درست آن در تعیین میزان دوز دارو در کنترل گلوکز تاثیر گذار خواهد بود. در این بخش انطباق ساختار کنترلی پیشنهاد شده با کنترل میزان قند خون در بیمار مبتلا به دیابت مطرح می‌شود. همانطور که بیان شد سطوح لغزش برای سیستم نرمال و مانده را می‌توان به صورت روابط (۲-۷) و (۳-۷) تعریف کرد:

$$s^N = \binom{2}{0} e_3^N + \binom{2}{1} \lambda^1 e_2^N + \binom{2}{2} \lambda^2 e_1^N \quad (۷-۲)$$

$$s^R = \binom{2}{0} R_3 + \binom{2}{1} \lambda^1 R_2 + \binom{2}{2} \lambda^2 R_1 \quad (۷-۳)$$

ورودی کنترلی u^N برای سیستم استفاده شده در این سیستم در معادله (۴-۷) بیان شده است.

$$u^N = \eta_N \text{sgn}(s^N) - \binom{2}{0} (f_N(\hat{X}^N) - x_d^{(n)}) - \binom{2}{1} \lambda^1 e_3^N - \binom{2}{2} \lambda^2 e_2^N \quad (۷-۴)$$

حالت‌های در نظر گرفته شده در این روش نیز سه بعدی بوده و میزان گلوکز خون، خطا (اختلاف میان مقدار مطلوب گلوکز و مقدار بدست آمده از گلوکز در حالتی که سیستم دچار عیب شده باشد) و میزان دامنه عیب وارد شده به سنسور سیستم در هر گام زمانی، می‌باشد.

در حالتی که سیستم دچار عیب شده است، حالت‌های سیستم با استفاده از روش شبکه‌های عصبی توابع پایه شعاعی مشاهده شده و با خروجی سیستم اصلی مقایسه می‌شوند. ورودی در این حالت که میزان دوز داروی تزریقی به بیمار می‌باشد u^f است که طبق رابطه (۵-۷) محاسبه می‌شود.

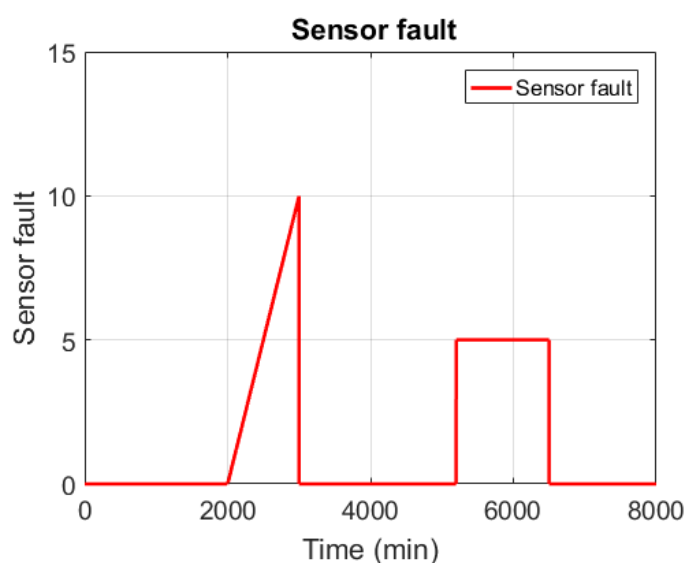
$$u^f = u^N + u^T \quad (۷-۵)$$

در رابطه (۶-۷)، u^T ورودی کنترلی تحمل‌پذیر بوده که در رابطه (۶-۷) نمایش داده شده است.

$$u^T = -u^R - (\eta_R \text{sgn}(s^R)) \quad (۷-۶)$$

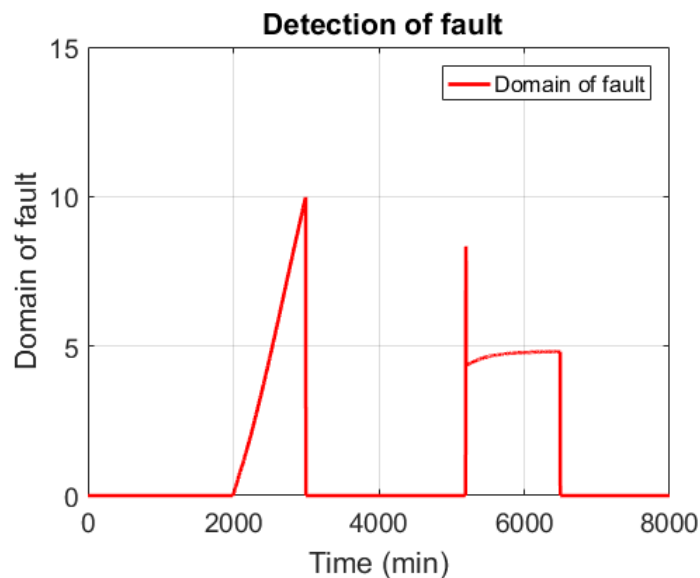
η_R مقداری ثابت و مثبت می‌باشد. u^R عملی می‌باشد که توسط عامل انتخاب شده است و به طور غیر مستقیم در تعیین میزان دوز داروی تزریقی به بیمار دخیل می‌باشد. شرایط اولیه که برای بیمار

در نظر گرفته شده است، بدین ترتیب می‌باشد: $G_0 = 250, X_0 = 10, I_0 = 20$. مقادیر مطلوب برای متغیرها که در نهایت می‌بایست به آن برسند، برای گلوکز ۸۵، مقدار عمومی انسولین ۱۱ و انسولین خون فرد ۳۴ می‌باشد. دوز داروی تزریقی که متغیر u^f می‌باشد، باید به گونه‌ای تعیین شود که هنگام بروز عیب در سیستم که در سنسور آن اتفاق می‌افتد، قادر به کنترل میزان گلوکز و نگه‌داشتن آن در سطح مطلوب باشد. عیوب وارد شده به سنسور سیستم در شکل (۶-۷) نمایش داده شده است.

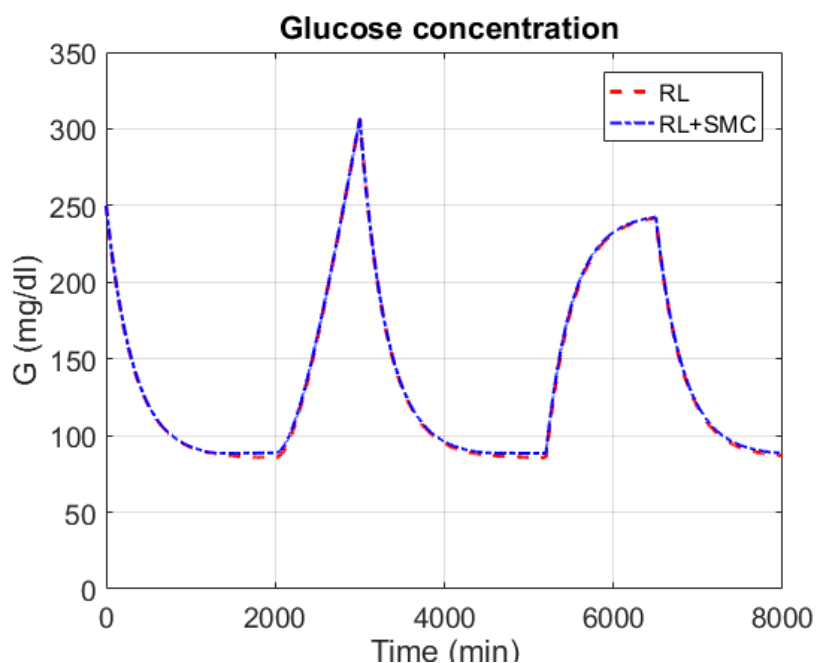


شکل ۶-۷: عیب وارد شده به سنسور سیستم

با توجه به عیب وارد شده به سنسور سیستم، تشخیص دامنه عیب توسط روش شبکه‌های عصبی در شکل (۷-۷) قابل مشاهده می‌باشد.

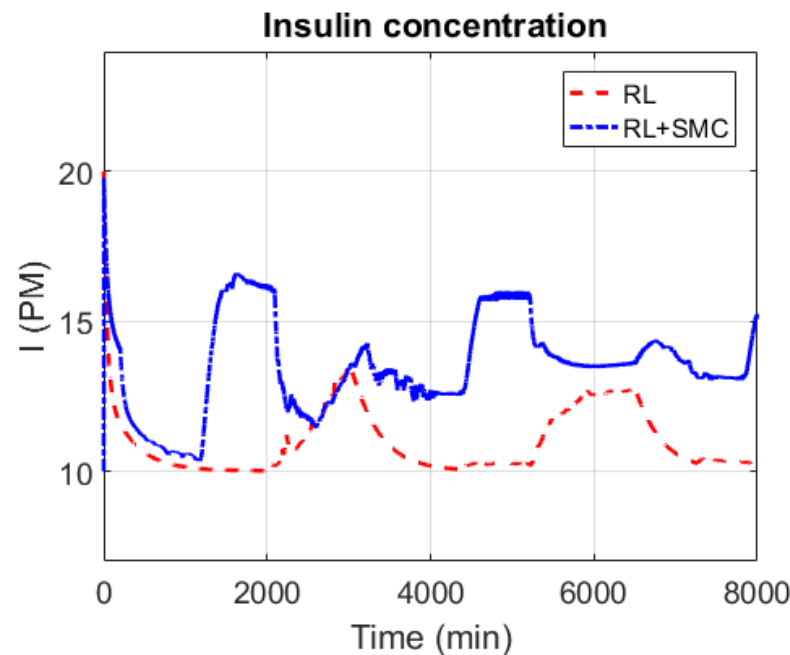


شکل ۷-۷: تشخیص دامنه عیب وارد شده به سنسور سیستم توسط روش شبکه‌های عصبی
تغییرات میزان گلوکز در حضور عیب وارد شده به سنسور سیستم توسط روش مسیرهای شایستگی و ترکیب روش مسیرهای شایستگی با کنترل مد لغزشی در شکل (۷-۸) قابل نمایش می‌باشد.



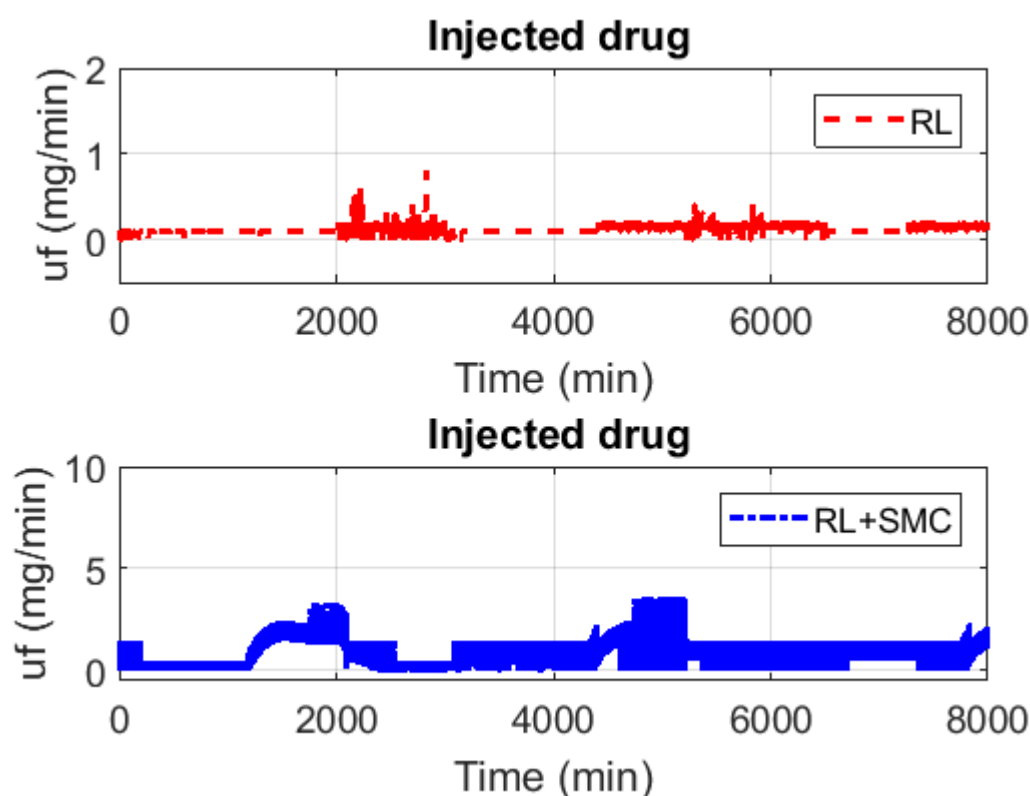
شکل ۷-۸: تغییرات میزان گلوکز در حضور عیب وارد شده به سنسور سیستم توسط روش مسیرهای شایستگی و ترکیب روش مسیرهای شایستگی و کنترل مد لغزشی ($RL+SMC$)
همانطور که در شکل فوق نشان داده شده است، هر دو روش پیشنهادی توانسته‌اند میزان گلوکز را در حضور عیب وارد شده به سنسور سیستم کنترل کرده و به مقدار مطلوب برسانند. اما روش پیشنهاد

شده در این مقاله علاوه بر سادگی روش با تزریق دوز داروی کمتر توانسته است کنترل میزان گلوکز و رساندن آن به مقدار مطلوب را انجام دهد که این امر باعث کاهش اثرات زیان بار دارو خواهد شد. میزان انسولین تزریق شده به بیمار نیز توسط هر دو روش ذکر شده در شکل (۷-۹) قابل نمایش می باشد.



شکل ۷-۹: میزان انسولین تزریق شده به بیمار توسط روش مسیرهای شایستگی و ترکیب روش مسیرهای شایستگی و کنترل مد لغزشی (RL+SMC)

با توجه به شکل (۷-۹)، میزان انسولین تزریق شده به بیمار در حالتی که از روش مسیرهای شایستگی به تنهایی استفاده شده است، علاوه بر اینکه میزان انسولین تزریق شده به بیمار کمتر می باشد، تزریق انسولین در این روش تطبیق پذیری بیشتر با عیب وارد شده به سنسور سیستم را دارا می باشد. همچنین این روش توانسته است بعد از اتمام عیب میزان انسولین را به مقدار مطلوب خود برساند. این در حالی می باشد که با ترکیب دو روش مسیرهای شایستگی و کنترل مد لغزشی میزان انسولین به درستی کنترل نشده و روی مقدار مطلوب قرار نمی گیرد. میزان دوز داروی تزریقی به بیمار در هر لحظه توسط هر دو روش در شکل (۷-۱۰) قابل مشاهده می باشد.



شکل ۷-۱۰: میزان دوز داروی تزریق شده به بیمار توسط مسیرهای شایستگی و ترکیب روش مسیرهای شایستگی و کنترل مد لغزشی (RL+SMC)

همانطور که در شکل (۷-۱۰) مشاهده می‌گردد، میزان دوز داروی تزریقی توسط روش مسیرهای شایستگی در هنگام بروز عیب در سیستم بسیار کمتر از حالت استفاده از ترکیب مسیرهای شایستگی و کنترل مد لغزشی می‌باشد. این امر باعث کاهش اثرات زیان‌بار دارو خواهد شد. مجموع دوز داروی تزریقی توسط هر دو روش در جدول (۷-۱) قابل نمایش می‌باشد.

جدول ۷-۱: مجموع دوز داروی تزریقی با استفاده از روش مسیرهای شایستگی و ترکیب روش مسیرهای شایستگی و کنترل مد لغزشی (RL+SMC)

روش مورد استفاده	میزان داروی تزریق شده
RL	1020.2mg/min
RL+SMC	6372.5mg/min

با توجه به جدول (۷-۱)، مجموع دوز داروی تزریق شده با استفاده از روش مسیرهای شایستگی به تنهایی بسیار کمتر نسبت به ترکیب روش مسیرهای شایستگی و کنترل مد لغزشی می‌باشد. همچنین در این روش داروی تزریقی به بیمار در لحظاتی که عیب به سیستم وارد شده است بیشتر از دیگر

لحظات می‌باشد. این در حالی می‌باشد که در حالت ترکیب این روش با کنترل مد لغزشی تزریق دارو تطبیقی با عیب وارد شده به سیستم نداشته است. این امر نشان می‌دهد که روش پیشنهادی علاوه بر سادگی کنترلر با تزریق دوز داروی کمتر و همچنین تطبیق پذیری بیشتر با عیب وارد شده به سنسور توانسته است میزان گلوکز را کنترل کرده و به مقدار مطلوب برساند.

۷-۴- شبیه‌سازی روش‌های یادگیری تقویتی در کنترل سرطان ملانوما و کنترل

تحمل‌پذیر عیب

در ابتدا برای مدل ریاضی سرطان ملانوما مطرح شده در فصل ۴ کنترل دوز دارو به کمک روش‌های یادگیری تقویتی را مطرح می‌کنیم سپس در مورد کنترل تحمل‌پذیر عیب صحبت خواهیم کرد. معادلات بیان‌کننده دینامیک بیماری ملانوما در روابط (۴-۷) تا (۴-۱۳) بیان شده‌اند. توضیحات کامل راجع به مدل در فصل ۴ آورده شده است. ثوابت سیستم در جدول (۷-۲) آورده شده‌اند.

جدول ۷-۲: مقادیر در نظر گرفته‌شده برای پارامترهای سیستم [۹۲]

پارامتر	معرفی پارامتر	مقدار
α_1	های CTL نرخ رشد	0.0065
α_2	های CTL نرخ کاهش	0.00038
α_3	های فعال OT1CTL ماکزیمم نرخ مرگ به وسیله	1×10^{-8}
α_4	های CTL نرخ حذف آنتی‌بادی‌ها در محفظه‌ی پوست به وسیله‌ی	0.0063
α_5	نرخ از هم پاشیده شدن آنتی‌بادی‌ها	0.0023
α_6	های فعال در محفظه‌ی پوست OT1CTL‌ها به CTL نرخ تغییر	0.0030
α_7	های فعال OT1CTL حداکثر نرخ تکرار	0.0028
α_8	های فعال در محفظه‌ی پوست و نقطه تزریق OT1CTL نرخ مرگ طبیعی	0.0014
α_9	نرخ مرگ آنتی‌بادی‌ها در محفظه‌ی تزریق	3.03×10^{-8}
α_{10}	سرعت از بین رفتن آنتی‌بادی در دو محفظه	0.001
α_{11}	سرعت حرکت از محفظه‌ی تزریق به محفظه‌ی پوست	0.00027
h	ها به وسیله غده تیموس CTL نرخ تزریق دوباره از	0.0003
K_1	آستانه تکرار آنتی‌بادی‌ها	10
K_2	مینیمم آستانه نرخ نابودی سلول‌های سرطانی	1
K_3	ماکزیمم آستانه نرخ نابودی سلول‌های سرطانی	50

196000	در موقعیت‌های بی‌خطر ها CTLتعداد	M
180000	مقدار اولیه سلول‌های سرطانی	C_0
760000	های فعال در محفظه‌ی تزریق CTL1 مقدار داروی تزریق شده به	P
1×10^6	مقدار داروی تزریق شده به آنتی‌بادی در محفظه‌ی تزریق	Q
1	تاخیر	τ

هدف تعیین میزان بهینه دوز دارو برای کاهش جمعیت سلول‌های سرطانی می‌باشد، به گونه‌ای که از اثرات زیان‌بار دارو بر روی سلول‌های سالم جلوگیری شود. در این مسئله محیط، مدل غیرخطی دارای تاخیر از یک بیمار مبتلا به سرطان ملانوما می‌باشد. عامل با جستجو در محیط و دریافت پاداش و جریمه یاد می‌گیرد، بهترین میزان دوز دارو را برای کاهش جمعیت سلول‌های سرطانی و کاهش اثرات زیان‌بار دارو را بدست آورد. برای این منظور از روش مسیرهای شایستگی در مسئله یادگیری تقویتی استفاده شده است.

۷-۴-۱- تعریف پارامترهای یادگیری تقویتی

حالت در نظر گرفته شده در این مقاله، جمعیت سلول‌های سرطانی می‌باشد که به صورت گسسته در ۲۰۰ بازه تعریف شده‌اند. بازه‌بندی از ۵۰۰۰ تا ۱۰۰۰۰۰۰ بدست آمده است. میزان داروهای تزریقی $(K_{in}(t, q), K_{in}(t, p))$ نیز به عنوان عمل در مسئله‌ی یادگیری تقویتی می‌باشند و هر دو بین بازه ۵۷۶۰۰۰ تا ۱۰۰۰۰۰۰ تقسیم‌بندی شده‌اند. عامل یاد می‌گیرد در هر لحظه بهترین میزان دوز دارو را تعیین کند که علاوه بر کاهش جمعیت سلول‌های سرطانی، از تاثیرات زیان‌بار دارو بر روی سلول‌های ایمنی بدن نیز جلوگیری شود. تعداد اعمال انتخابی توسط عامل ۱۶۱ عدد می‌باشد. سیاست در نظر گرفته شده برای انتخاب اعمال $\epsilon - greedy$ می‌باشد. تابع پاداش در مسئله یادگیری تقویتی با توجه به هدف مورد نظر در هر مسئله تعیین می‌گردد و برخلاف روش‌های کنترل بهینه از هیچ رابطه ریاضی به‌خصوصی تبعیت نمی‌کند. در این مقاله با توجه به اینکه هدف کاهش جمعیت سلول‌های سرطانی و افزایش ایمنی بدن با تزریق بهترین میزان دوز دارو، می‌باشد، تابع پاداش بر اساس تغییرات میزان سلول‌های سرطانی بدن و همچنین تغییرات هر دو نوع دارو در نظر گرفته شده است. با این تعریف

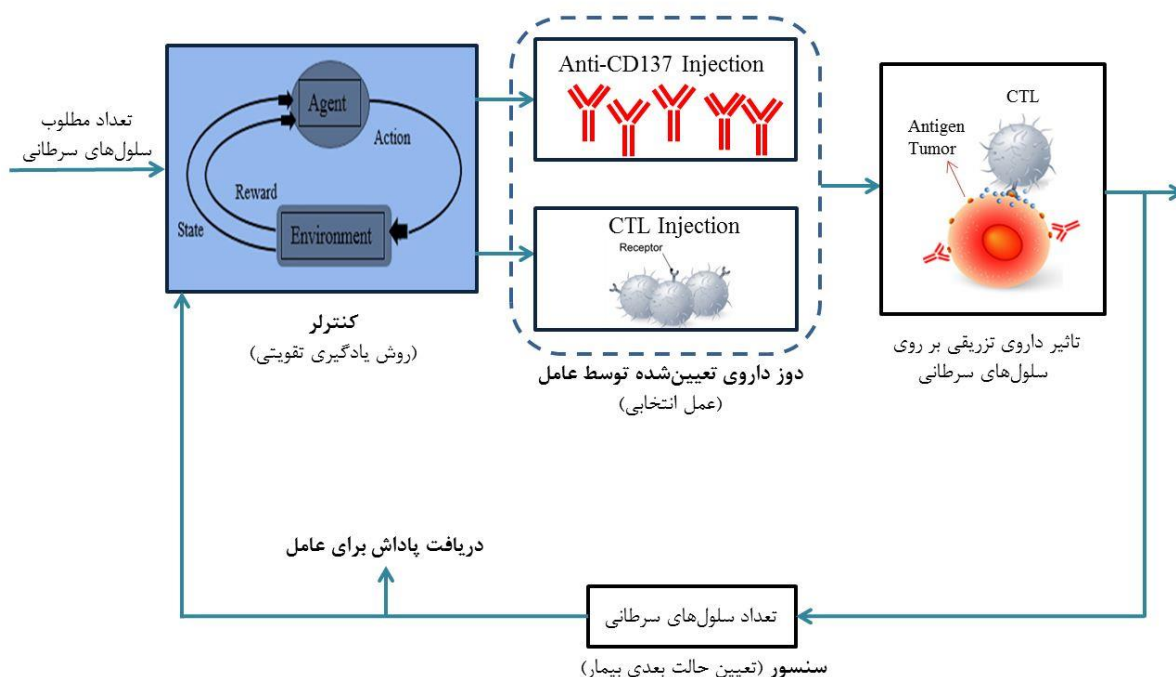
برای تابع پاداش، از اثرات زیان بار دارو بر روی سلول‌های سالم نیز جلوگیری خواهد شد. تابع پاداش به صورت رابطه (۷-۷) تعریف شده است.

$$Reward = -\log\left(\frac{C_{new}}{C_{old}}\right) - (dose1_{new} - dose1_{old}) - (dose2_{new} - dose2_{old}) \quad (7-7)$$

در رابطه (۷-۱۴)، C_{old} جمعیت سلول‌های سرطانی بدن در لحظه‌ی قبل، C_{new} جمعیت سلول‌های سرطانی بدن در لحظه‌ی فعلی، می‌باشند. $(dose1_{new} - dose1_{old})$ ، بیان‌کننده میزان اختلاف داروی تزریقی نوع اول در لحظه فعلی و لحظه قبل می‌باشد. $(dose2_{new} - dose2_{old})$ ، بیان‌کننده میزان اختلاف داروی تزریقی نوع دوم در لحظه فعلی و لحظه قبل می‌باشد. بدین ترتیب پاداش دریافتی مثبت خواهد بود. روند عملکرد پارامترهای یادگیری تقویتی در کنترل جمعیت سلول‌های سرطانی در شکل (۷-۱۱) قابل نمایش می‌باشد.

با توجه به شکل (۷-۱۱) و همان‌طور که ذکر شد، پارامترهای یادگیری تقویتی حالت، عمل و پاداش دریافتی می‌باشند. حالت در این مقاله، حالت‌های مختلف بیمار برای مقادیر مختلف سلول‌های سرطانی در هر لحظه می‌باشد. عملی که عامل در هر لحظه انتخاب می‌کند، میزان دوز داروی تزریقی می‌باشد که همان‌طور که بیان شد، در این مقاله دو نوع داروی تزریقی برای کاهش جمعیت سلول‌های سرطانی در نظر گرفته شده است. روند کاهش سلول‌های سرطانی با توجه به آنچه که در شکل (۷-۱۱) نشان داده شده است، بدین صورت می‌باشد که ابتدا در مرحله یادگیری، جدولی تحت عنوان "جدول ارزش‌گذاری بر جفت حالت و عمل (جدول Q)" تعریف می‌گردد. این جدول با مقدار صفر مقداردهی اولیه می‌شود. مقادیر این جدول در روند اجرای الگوریتم توسط رابطه (۷-۸) مقداردهی شده و بروزرسانی می‌گردند.

$$Q(s, a) = Q(s, a) + \alpha \delta e(s, a) \quad (7-8)$$



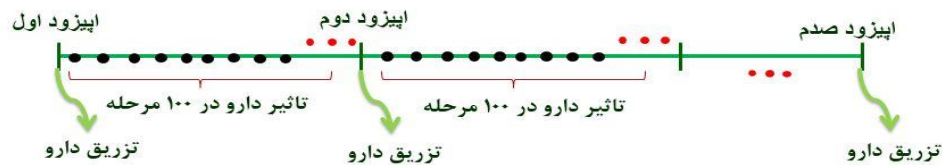
شکل ۷-۱۱: روند عملکرد پارامترهای یادگیری تقویتی در کنترل جمعیت سلول‌های سرطانی

عامل در هر لحظه با تاثیرپذیری از این جدول عمل انتخابی که میزان دوز داروی تزریقی می‌باشد را انتخاب می‌کند. این دارو به مدل غیرخطی از بیمار مبتلا به سرطان ملانوما اعمال می‌گردد. حالت بعدی بیمار (جمعیت سلول‌های سرطانی) محاسبه می‌گردد. در این گام عامل به خاطر عمل انتخابی که چه میزان باعث کاهش یا افزایش سلول‌های سرطانی شده است، از محیط پاداش دریافت می‌کند. در این مقاله همانطور که بیان شد و طبق رابطه (۱۷) پاداش دریافتی بر اساس تغییرات سلول‌های سرطانی و میزان دوز داروی تزریقی در نظر گرفته شده است. این انتخاب به این دلیل می‌باشد که علاوه بر کاهش سلول‌های سرطانی میزان دوز داروی تزریقی نیز کاهش پیدا کند تا از اثرات زیان‌بار دارو بر روی سلول‌های سالم نیز جلوگیری گردد. این اعمال در چندین گام زمانی تکرار می‌گردند. برورسانی جدول Q بر عمل انتخابی توسط عامل تاثیر گذاشته و باعث بهبود آن می‌گردد.

• تاثیر نیمه عمر دارو در بدن

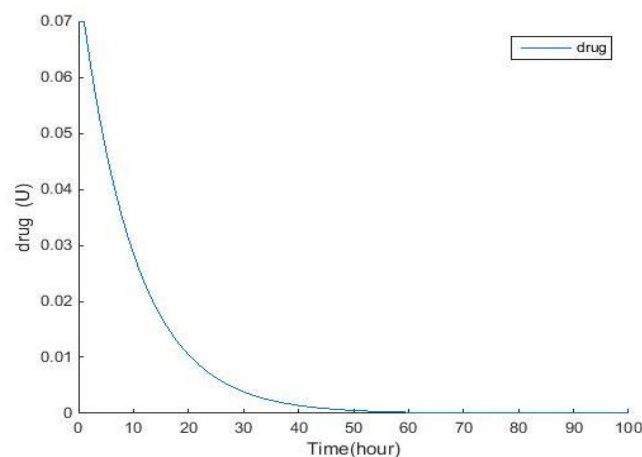
داروی مورد نظر در این رساله برای درمان سرطان ملانوما اینترفرون می‌باشد که دارای نیمه عمر ۱۰ ساعت می‌باشد و با توجه به این مورد تزریق دارو در هر ۱۰ ساعت انجام گرفته است. هر ۶ دقیقه

یکبار میزان داروی موجود در بدن بر طبق روابط بدست آمده که در واقع نشان دهنده ۱۰۰ گام می‌باشد. در شکل (۷-۱۲)، این گام‌های زمانی قابل مشاهده می‌باشند.



شکل ۷-۱۲: نمایش گام‌های زمانی تزریق دارو

با توجه به شکل فوق، تزریق دارو در ابتدای هر اپیزود انجام می‌گیرد و تاثیر آن تا انتهای اپیزود با توجه به نیمه عمر دارو لحاظ شده است. شکل (۷-۱۳)، نیز روند کاهشی میزان دارو در بدن بیمار در طول ۱۰۰ ساعت را نشان می‌دهد.



شکل ۷-۱۳: روند کاهش میزان دارو در بدن در طول ۱۰۰ ساعت (اثر نیمه عمر داروی اینترفرون)

رابطه (۷-۹)، غلظت پلاسمای دارو را نشان می‌دهد که مقدار اولیه آن برابر با صفر خواهد بود و

طبق رابطه (۷-۱۰) محاسبه می‌شود. [۹۷]

$$C_n(t) = \frac{D}{V} \cdot \frac{(1 - e^{-nK\tau})}{(1 - e^{-K\tau})} \cdot e^{-Ket} \quad (۷-۹)$$

$$C_0 = \frac{D}{V} \quad (۷-۱۰)$$

در روابط فوق، t زمان نمونه برداری، n تعداد گام‌های زمانی برای مشاهده تاثیر دارو (در اینجا ۱۰۰

گام زمانی در نظر گرفته شده است) و τ بازه‌ی تزریق دارو می‌باشد که با توجه به نیمه عمر داروی

اینترفرون ۱۰ ساعت لحاظ شده است. D نشان دهنده میزان دوز داروی در هر گام است و طبق رابطه

(۱۱-۷) این مقدار محاسبه می‌شود، که در این رابطه، D_0 میزان دوز داروی تزریقی در ابتدای هر ده ساعت می‌باشد.

$$D = D_0 \cdot e^{-K_e t} \quad (۱۱-۷)$$

در رابطه (۱۱-۷)، K_e نشان دهنده نرخ حذف دارو می‌باشد که این مقدار طبق رابطه (۱۲-۷) محاسبه می‌شود و مقدار اولیه آن برابر با یک در نظر گرفته شده است.

$$K_e = \frac{CL}{V} \quad (۱۲-۷)$$

در رابطه (۱۲-۷)، CL سرعت حذف دارو می‌باشد که مقدار آن بر اساس معادله زیر محاسبه می‌شود.

$$CL = K_e \times V \quad (۱۳-۷)$$

در روابط فوق V حجم توزیع دارو در بدن می‌باشد و طبق رابطه (۱۴-۷) محاسبه می‌شود. که در ابتدای کار ۱۰۰۰۰ در نظر گرفته شده است.

$$V = \frac{D}{K_e \cdot T} \cdot \frac{(1 - e^{-K_e T})}{[C_{max} - (C_{min} \cdot e^{-K_e T})]} \quad (۱۴-۷)$$

در رابطه فوق، T معرف زمان از بین رفتن دارو می‌باشد. C_{min} و C_{max} به ترتیب ماکزیمم و مینیمم غلظت پلاسمای دارو در هر لحظه می‌باشند که طبق رابطه (۱۵-۷) و (۱۶-۷) محاسبه می‌شوند. مقدار مطلوب غلظت پلاسما بین ۱۰ تا ۲۰ میکروگرم می‌باشد.

$$C_{max} = \frac{D}{V \cdot K_e \cdot T} \cdot \frac{(1 - e^{-K_e T})}{(1 - e^{-K_e \tau})} \quad (۱۵-۷)$$

$$C_{min} = C_{max} \cdot e^{K_e(\tau - T)} \quad (۱۶-۷)$$

۷-۴-۲ شبیه‌سازی روش یادگیری-Q برای کنترل سرطان ملانوما

در این بخش برای اینکه عملکرد روش پیشنهادی در افزایش سرعت کاهش سلول‌های سرطانی مورد بررسی قرار گیرد، از یکی از روش‌های کنترل کلاسیک (کنترل بهینه) و یکی از الگوریتم‌های روش

یادگیری تقویتی به نام یادگیری Q جهت مقایسه با روش انتخابی استفاده شده است.

الگوریتم یادگیری Q روش یکی از الگوریتم‌های مرسوم در روش یادگیری تفاوت گذرا می‌باشد که در آن به‌روزرسانی ارزش جفت حالت و عمل در همان لحظه انجام می‌شود و نیازی به اتمام فرایند جهت ارزش‌گذاری زوج‌های حالت و عمل نمی‌باشد. به‌روزرسانی ارزش جفت حالت و عمل (s_t, a_t) طبق

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)] \quad (۱۷-۷)$$

معادله (۱۷-۷) محاسبه می‌شود.

در رابطه (۲۴-۷)، α نرخ یادگیری و γ که نرخ فراموشی می‌باشد. لازم به ذکر می‌باشد تابع پاداش، نحوه انتخاب اعمال و پارامترهای نرخ یادگیری و نرخ فراموشی در هر دو روش مسیرهای شایستگی و یادگیری Q یکسان در نظر گرفته شده است.

در بخش کنترل بهینه از روش "کنترل کننده بهینه رگولاتور درجه دوم خطی"^۱ با نام اختصاری LQR برای مقایسه با روش انتخابی استفاده شده است. این روش برای سیستم‌های خطی مورد استفاده قرار می‌گیرد. هدف اصلی در این روش یافتن ورودی‌های کنترلی u^* است که با اعمال آن به سیستم خطی تعریف شده در رابطه (۱۸-۷)، علاوه بر اینکه سیستم پایدار می‌شود و قیود تعریف شده برای آن برقرار می‌گردد، تابع هزینه تعریف شده نیز می‌بایست مینیمم گردد. در این روش متغیرهای حالت نیز با کمترین تلاش کنترلی به سمت صفر همگرا می‌شوند [۹۸].

$$\dot{x}(t) = A(x)x(t) + B(x)u(t) \quad (۷-۱۸)$$

در این روش تابع هزینه بر اساس رابطه (۱۹-۷) تعریف می‌گردد [۹۸].

$$J(x, u, t) = \frac{1}{2}x^T(t_f)Hx(t_f) + \int_{t_0}^{t_f} \left\{ \frac{1}{2}x^T(t)Qx(t) + \frac{1}{2}u^T(t)Ru(t) \right\} dt \quad (۷-۱۹)$$

¹ Linear Quadratic Regulator

در رابطه فوق، H و Q ماتریس‌های وزنی متقارن مثبت می‌باشند و R نیز ماتریس وزنی اکیدا مثبت می‌باشد. رابطه (۷-۲۰) فرم ورودی کنترلی را نشان می‌دهد.

$$u^* = -R^{-1}BT(x)Kx \quad (7-20)$$

در رابطه (۷-۲۰)، K ضریب لاگرانژ می‌باشد و از حل یک معادله ریکاتی بدست می‌آید. این معادله در رابطه (۷-۲۱) نشان داده شده است.

$$k = -KA - A^TK - Q + KBR^{-1}B^TK \quad (7-21)$$

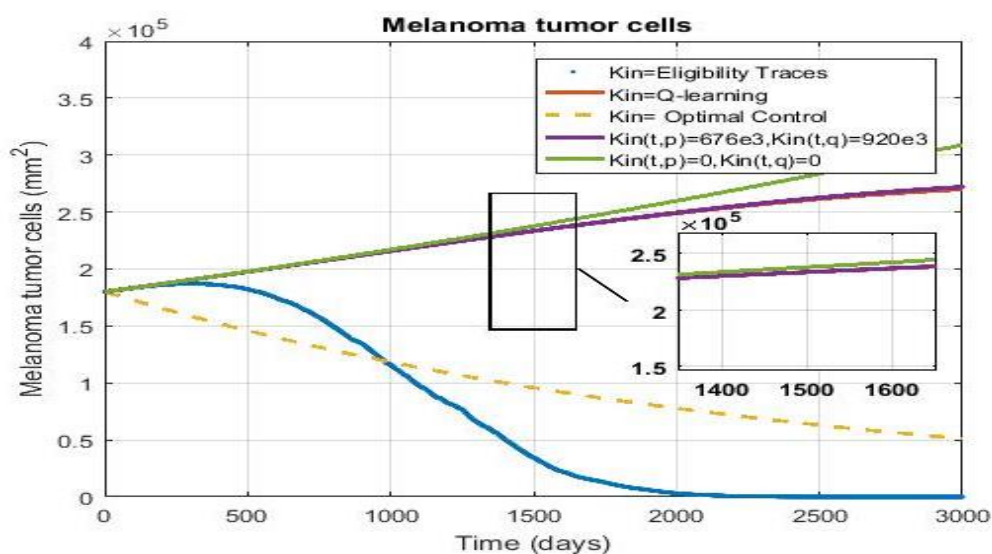
برای سیستم مورد استفاده در این مقاله که یک مدل غیر خطی از بیمار مبتلا به سرطان ملانوما می‌باشد، ابتدا خطی‌سازی انجام شده است و سپس کنترلر مربوطه با استفاده از روش LQR برای آن طراحی شده است. در نهایت با استفاده از ماتریس‌های بدست آمده، کنترلر مربوطه برای مدل غیرخطی ذکر شده طراحی شده است. شرایط اولیه برای یک بیمار مبتلا به ملانوما در جدول (۷-۳) آورده شده است.

جدول ۷-۳: شرایط اولیه برای یک بیمار مبتلا به ملانوما

متغیر	مقدار اولیه	متغیر	مقدار اولیه
$C_0(0)$	$18 \times 10^4 (mm^2)$	$Ab_0(0)$	$0 (mg/ml)$
$E(0)$	$0 (Cell/Kg)$	$A_0(0)$	$0 (\mu g \text{ of protein})$
$A_s(0)$	$0 (mg/ml)$	$N_0(0)$	$196 \times 10^3 (cell/days)$

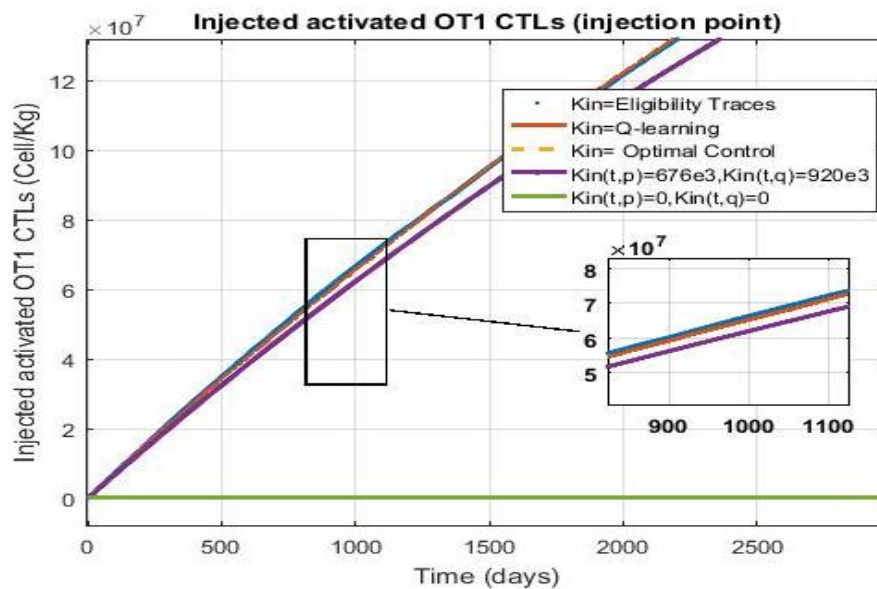
برای اینکه عملکرد روش مسیرهای شایستگی بهتر نشان داده شود، رفتار سیستم در حالت تعیین دارو با استفاده از روش مسیرهای شایستگی علاوه بر روش یادگیری Q و کنترل بهینه با تزریق دوز داروی

صفر و مقدار ثابت در هر لحظه، مقایسه می‌شود. مقدار ثابت برای هر دارو، مقادیر $K_{in}(t, p) = 676 \times$ برای $OT1CTL$ ها و $K_{in}(t, q) = 920 \times 10^3 (mg/ml)$ برای آنتی‌بادی‌ها در نظر گرفته شده است. شکل (۷-۱۴) جمعیت سلول‌های سرطانی را تزریق در حالت تعیین دوز دارو با استفاده روش‌های مسیرهای شایستگی، یادگیری Q ، کنترل بهینه، دوز داروی ثابت و دوز صفر را نشان می‌دهد.



شکل ۷-۱۴: رفتار جمعیت سلول‌های سرطانی در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q ، کنترل بهینه، مقدار ثابت و دوز داروی صفر همانطور که در شکل فوق نشان داده شده است، جمعیت سلول‌های سرطانی در حالت تعیین دوز دارو با استفاده از روش مسیرهای شایستگی در مدت زمان درمان در نظر گرفته شده کاهش سریع‌تری نسبت به دیگر روش‌ها داشته است. تزریق دوز دارو در روش انتخابی نیز کمتر بوده که این امر کاهش اثرات زیانبار دارو را به همراه داشت. این در حالی می‌باشد که دیگر روش‌ها قادر به از بین بردن سلول‌های سرطانی به‌طور کامل در این مدت درمان نبوده‌اند. شکل (۷-۱۵) رفتار $OT1CTL$ های فعال در محفظه‌ی نقطه‌ی تزریق در حالت تعیین دوز دارو با استفاده از پنج حالت ذکر شده را نشان می‌دهد. $OT1CTL$ ها به عنوان سلول‌های ایمنی بدن و کشنده سلول‌های سرطانی می‌باشند. بنابراین تا زمانی که سلول‌های سرطانی در بدن از بین نرفته‌اند، مقدار آن‌ها افزایش پیدا کرده و در نهایت رشد

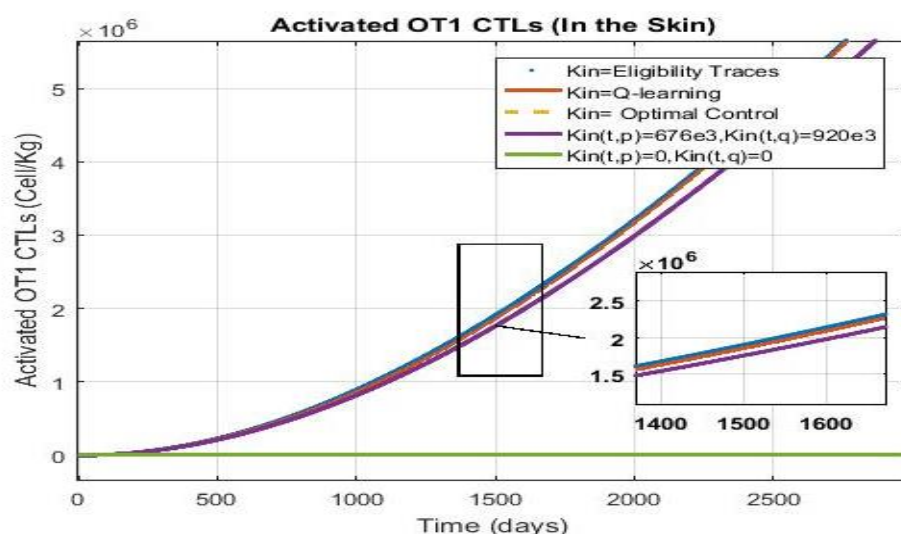
آنها متوقف شده و به مقدار ثابت میرسند. همان طور که در شکل (۷-۱۵) نشان داده شده است، در حالت تزریق دارو برای کاهش سلول های سرطانی OT1CTL ها افزایش پیدا می کنند. این امر در حالت استفاده از روش های مسیرهای شایستگی و الگوریتم یادگیری Q با سرعت بیشتری نسبت به تزریق دوز داروی ثابت اتفاق افتاده است. در حالت استفاده از روش کنترل بهینه نیز سرعت افزایش OT1CTL ها در مدت زمان در نظر گرفته شده بسیار پایین بوده و نزدیک به صفر می باشد.



شکل ۷-۱۵: رفتار OT1CTL های فعال تزریق شده به نقطه تزریق در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل بهینه، مقدار ثابت و دوز داروی صفر

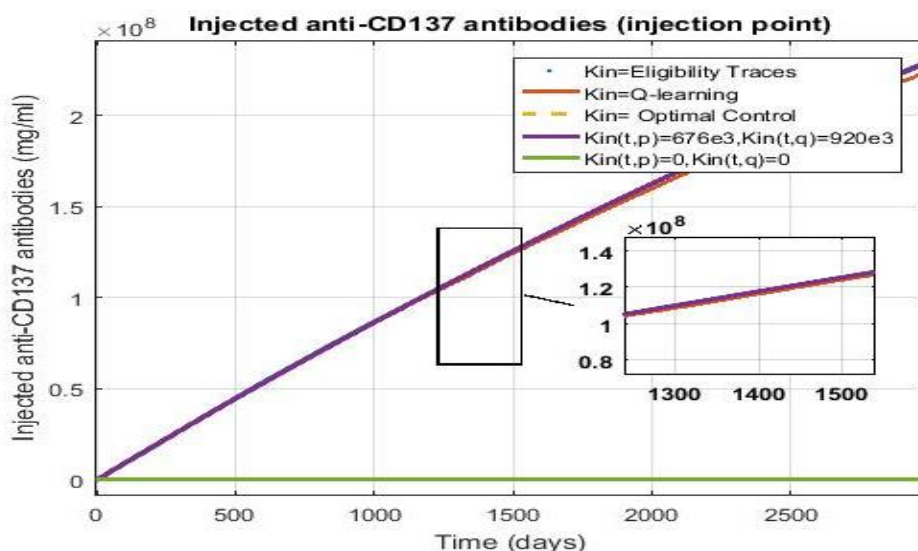
شکل (۷-۱۶) رفتار OT1CTL های فعال در محفظه پوست در حالت تعیین میزان دوز دارو با استفاده از روش های مسیرهای شایستگی، یادگیری Q، کنترل بهینه و دوز داروی ثابت و دوز صفر را نشان می دهد. در این شکل نیز، OT1CTL های که در محفظه پوست می باشند، برای از بین بردن سلول های سرطانی در این ناحیه افزایش می یابند. در حالت استفاده از روش های مسیرهای شایستگی و الگوریتم یادگیری Q این متغیر نیز با سرعت بیشتری نسبت به تزریق دوز داروی ثابت افزایش یافته است. بدون تزریق دارو نیز تغییراتی در این متغیر دیده نمی شود. در حالت استفاده از روش کنترل بهینه نیز سرعت افزایش OT1CTL ها در محفظه پوست و در مدت زمان در نظر گرفته شده بسیار پایین بوده و

نزدیک به صفر می‌باشد.



شکل ۷-۱۶: رفتار OT1 CTL های فعال در محفظه پوست، در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل بهینه مقدار ثابت و دوز داروی صفر

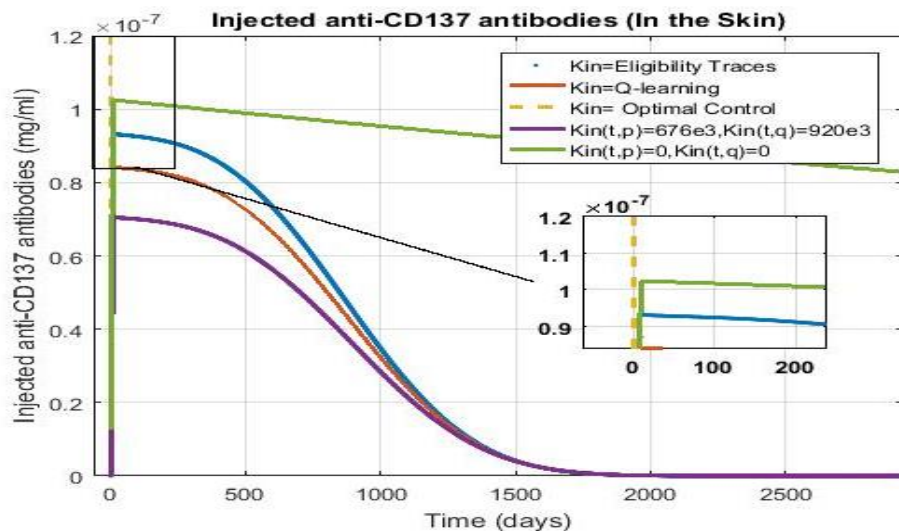
شکل (۷-۱۷) تغییرات میزان آنتی بادی CD137 تزریق شده در محفظه‌ی نقطه‌ی تزریق در کل مدت درمان و با استفاده از هر پنج روش ذکر شده را نشان می‌دهد.



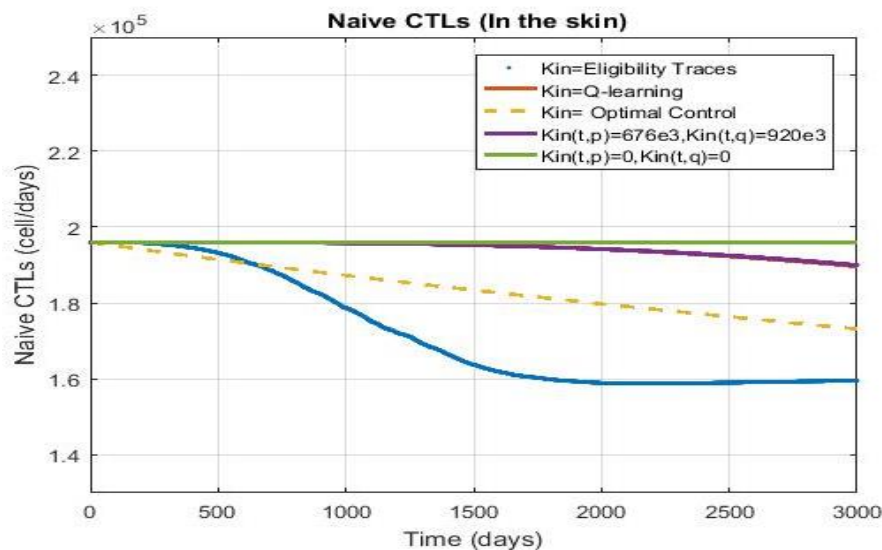
شکل ۷-۱۷: تغییرات آنتی‌بادی CD137 تزریق شده به نقطه تزریق در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل بهینه، مقدار ثابت و دوز داروی صفر

همان‌طور که در شکل فوق نشان داده شده است، میزان آنتی بادی‌های ترشح شده در محفظه نقطه تزریق نیز برای کاهش سلول‌های سرطانی افزایش پیدا می‌کنند. سرعت افزایش آن‌ها در حالت

استفاده از روش مسیرهای شایستگی بیشتر از روش الگوریتم یادگیری Q می‌باشد. در حالت استفاده از روش کنترل بهینه نیز سرعت تغییرات بسیار پایین بوده و نزدیک به صفر می‌باشد. در شکل (۷-۱۸) تغییرات میزان آنتی‌بادی‌ها در محفظه‌ی پوست در حالت تعیین دوز دارو با استفاده از هر پنج روش ذکر شده قابل نمایش می‌باشد.

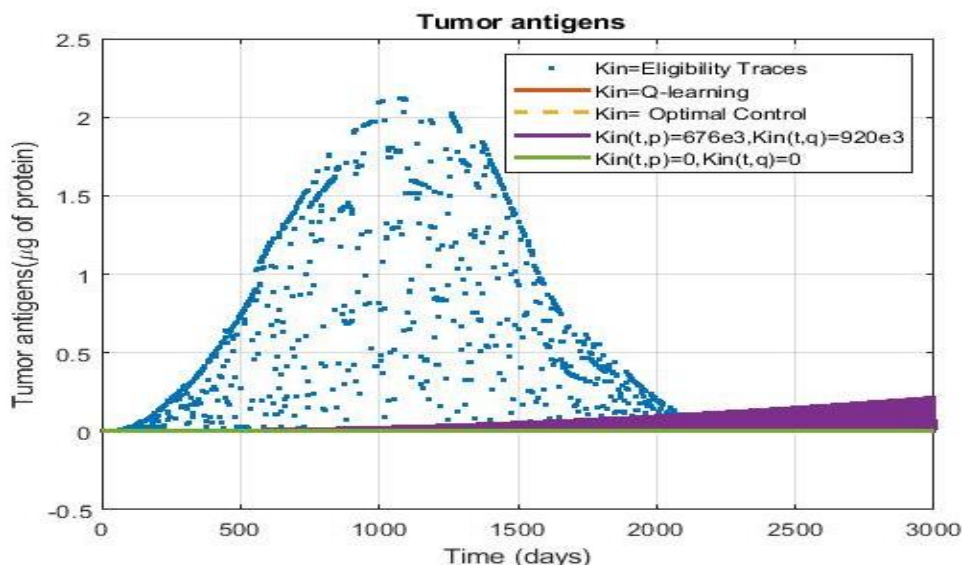


شکل ۷-۱۸: تغییرات آنتی‌بادی CD137 در محفظه پوست، در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل، مقدار ثابت و دوز داروی صفر با توجه به شکل فوق، آنتی‌بادی‌ها در محفظه پوست در ابتدا که جمعیت سلول‌های سرطانی بیشتر می‌باشند افزایش می‌یابند و با کاهش سلول‌های سرطانی میزان آن‌ها کاهش پیدا کرده و از بین می‌روند. همانطور که در این شکل نشان داده شده است، آنتی‌بادی‌ها در محفظه پوست در حالت استفاده از روش کنترل بهینه افزایش بسیار زیادی داشته است. شکل (۷-۱۹)، رفتار CTL‌های ساده با استفاده از هر پنج روش ذکر شده را نشان می‌دهد.



شکل ۷-۱۹: رفتار CTL های ساده، در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل بهینه، مقدار ثابت و دوز داروی صفر

CTL های ساده در بدن هر فرد به مقدارهای متفاوتی وجود دارد. با وارد شدن سلول های سرطانی به بدن، این CTL ها تبدیل به CTL فعال و کشته می شوند، برای اینکه سلول های سرطانی را از بین ببرند. بنابراین CTL های ساده با حذف سلول های سرطانی کاهش پیدا کرده و سپس افزایش می یابند. همان طور که در شکل (۷-۱۹) نشان داده شده است، در حالت استفاده از روش مسیرهای شایستگی طبق روند ذکر شده، ابتدا CTL های ساده کاهش پیدا کرده اند و با حذف سلول های سرطانی مقدار آنها ثابت می شود. این در حالی می باشد که در مدت زمان در نظر گرفته شده در حالت استفاده از روش تزریق دوز داروی ثابت، الگوریتم یادگیری Q، کنترل بهینه، CTL های ساده فقط کاهش پیدا کرده اند. کاهش CTL های ساده به معنای تبدیل آن ها به CTL های فعال می باشد. کاهش سریع CTL های ساده باعث تولید CTL های فعال بیشتر می شود. در حالت بدون تزریق دارو نیز مقدار این CTL ها تغییری نکرده است. بنابراین روش مسیرهای شایستگی بهتر توانسته است میزان CTL ها را کنترل کند. شکل (۷-۲۰)، تغییرات میزان آنتی ژن تومور را در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q، کنترل بهینه، دوز داروی ثابت و بدون تزریق دارو نشان می دهد.



شکل ۷-۲۰: تغییرات آنتی‌ژن تومور، در حالت استفاده از روش مسیرهای شایستگی، الگوریتم یادگیری Q ، کنترل بهینه، مقدار ثابت و دوز داروی صفر

همان‌طور که در شکل فوق نشان داده شده است، میزان آنتی‌ژن‌های تومور در حالت استفاده از روش مسیرهای شایستگی با سرعت بیشتری نسبت به دیگر روش‌ها کاهش پیدا کرده‌اند. دلیل این امر، استفاده از سلول‌های سرطانی به صورت رابطه $(-\log(\frac{y_{new}}{y_{old}}))$ در تابع پاداش می‌باشد. که با تاثیر مستقیم سلول‌های سرطانی در تغییرات تومور آنتی‌ژن‌ها کاهش آن‌ها با سرعت بالا اتفاق افتاده است. این امر در حالت استفاده از الگوریتم یادگیری Q و کنترل بهینه اتفاق نیفتاده است. لازم به ذکر می‌باشد که به دلیل اینکه روند تغییرات آنتی‌ژن تومور بهتر نشان داده شود، این شکل با تعداد گام زمانی بیشتر رسم شده است.

میزان دوز داروی تزریقی توسط هر چهار روش در جدول (۷-۴) آورده شده است. بر اساس این جدول مجموع دوز داروی تزریقی با استفاده از روش مسیرهای شایستگی کمتر از روش‌های الگوریتم یادگیری Q ، کنترل بهینه و روش تزریق دوز داروی ثابت می‌باشد که باعث کاهش اثرات زیان‌بار دارو خواهد شد. همان‌طور که در شکل (۷-۱۴) نشان داده شد، سلول‌های سرطانی در حالت استفاده از روش مسیرهای شایستگی با سرعت بیشتر و در تعداد گام کمتر نسبت به دیگر روش‌ها از بین رفته‌اند. علاوه بر این همان‌طور که در جدول (۷-۴) نشان داده شده است، مجموع دوز داروی تزریقی به بیمار

در کل مدت درمان نیز در حالت استفاده از این روش کمتر می‌باشد که این امر باعث کاهش اثرات زیان‌بار دارو و آسیب رسیدن به سلول‌های سالم خواهد شد.

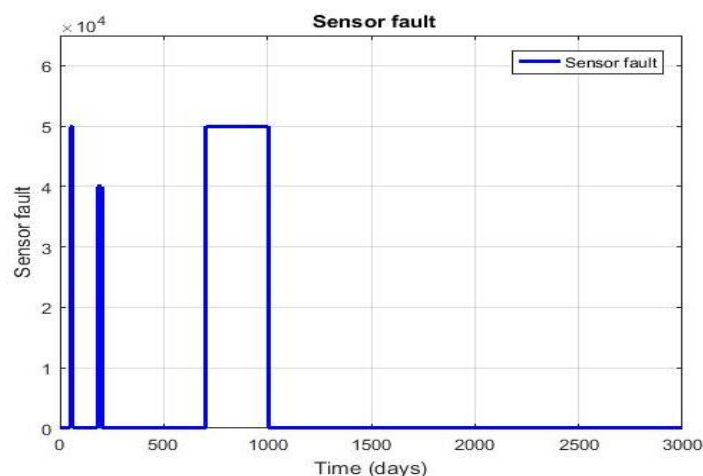
جدول ۴-۷: مقایسه مجموع داروهای تزریقی به بیمار در سه حالت ثابت، الگوریتم یادگیری Q ، مسیرهای شایستگی و با استفاده از روش کنترل بهینه

روش تزریق دارو	مجموع داروی تزریقی به بیمار تا زمان از بین رفتن سلول‌های سرطانی
تزریق داروی ثابت	$K_{in}(t, p) = 9.464 \times 10^9 \text{ Cell/Kg}$
	$K_{in}(t, q) = 1.288 \times 10^{10} \text{ mg/ml}$
دارو با استفاده از روش الگوریتم تزریق Q یادگیری	$K_{in}(t, p) = 1.4318 \times 10^{10} \text{ Cell/Kg}$
	$K_{in}(t, q) = 1.8269 \times 10^{10} \text{ mg/ml}$
دارو با استفاده از روش مسیر شایستگی	$K_{in}(t, p) = 1.7958 \times 10^9 \text{ Cell/Kg}$
	$K_{in}(t, q) = 2.3017 \times 10^9 \text{ mg/ml}$
کنترل بهینه	$K_{in}(t, p) = 2,2344 \times 10^9 \text{ Cell/Kg}$
	$K_{in}(t, q) = 1.0847 \times 10^{13} \text{ mg/ml}$

۷-۴-۳- کنترل جمعیت سلول‌های سرطانی در حضور عیب سنسور با استفاده از مسیرهای

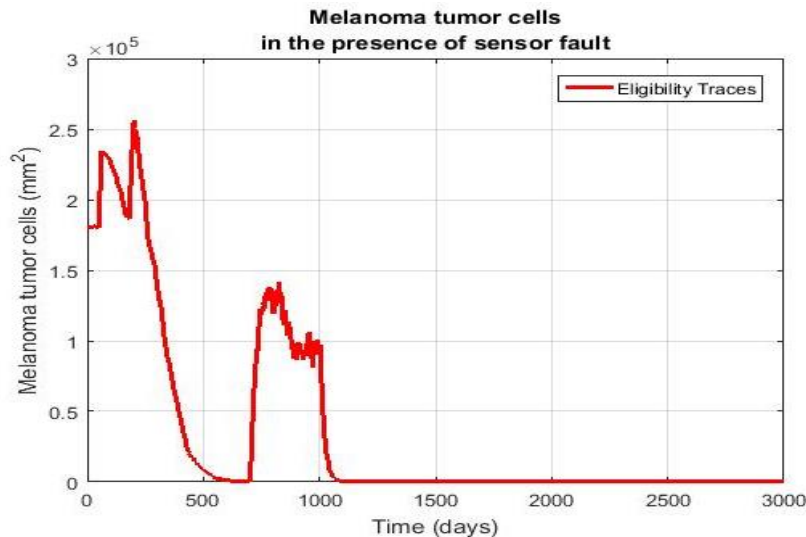
شایستگی

در این بخش عملکرد کنترل‌کننده پیشنهادی (روش مسیرهای شایستگی) در حضور عیب در سنسور سیستم مورد بررسی قرار گرفته است. این امر منطبق مفاهیم فصل ۷ در تشخیص خطا انجام گرفته است. عیب وارد به سنسور سیستم در شکل (۷-۲۱) نشان داده شده است.

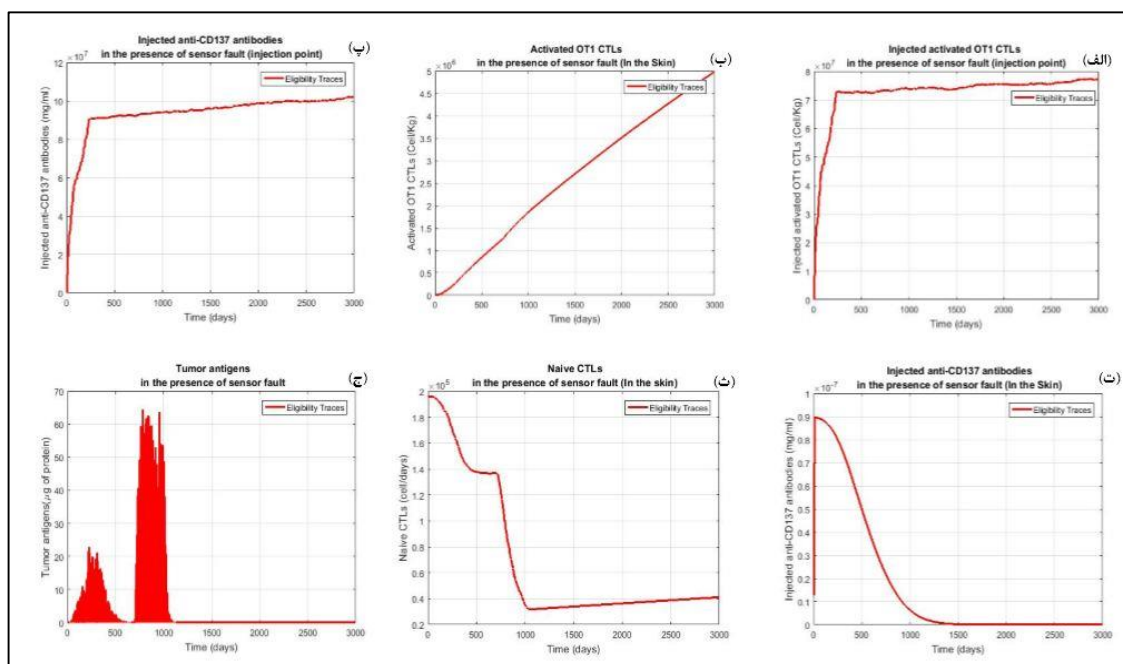


شکل ۷-۲۱: عیب وارد شده به سنسور سیستم

عملکرد کنترلر پیشنهادی (روش مسیرهای شایستگی) در کاهش جمعیت سلول‌های سرطانی در حضور عیب در سنسور سیستم در شکل (۷-۲۲) نشان داده شده است.



شکل ۷-۲۲: رفتار جمعیت سلول‌های سرطانی در حالت استفاده از روش مسیرهای شایستگی همانطور که در شکل (۷-۲۲) نشان داده شده است، با رخ دادن عیب در سنسور سیستم باز هم کنترلر پیشنهادی توانسته است جمعیت سلول‌های سرطانی را کنترل کرده و آن‌ها را از بین ببرد. رفتار دیگر متغیرها نیز در حضور عیب در سنسور سیستم کنترل شده می‌باشد و در شکل (۷-۲۳) آورده شده‌اند.



شکل ۷-۲۳: رفتار متغیرهای سیستم در حضور عیب سنسور

همانطور که در این شکل نشان داده شده است، در نمودار (الف)، CTLهای تزریق شده به محفظه نقطه تزریق تا زمان از بین نرفتن کامل سلولهای سرطانی افزایش یافته است سپس به یک مقدار ثابت رسیده است. در نمودار (ب) نیز CTLهای محفظه پوست برای کاهش سلولهای سرطانی افزایش پیدا کرده‌اند. در نمودار (پ)، آنتی‌بادی نوع CD-137 تزریق شده به نقطه تزریق تا زمان کاهش سلولهای سرطانی به بیمار تزریق شده و در نهایت به مقدار ثابت رسیده است. در نمودار (ت) نیز، آنتی‌بادی نوع CD-137 تزریق شده به محفظه پوست ابتدا افزایش پیدا کرده و با کاهش سلولهای سرطانی از بین رفته است. CTLهای ساده در نمودار (ث) نیز با تبدیل شدن به CTL کاهش پیدا کرده و یک مقدار ثابت می‌رسند. در نمودار (ج) آنتی‌ژن توموری نیز در هنگام اعمال عیب افزایش پیدا کرده است. مجموع میزان دوز داروی تزریقی در حالت اعمال عیب به سنسور بدین صورت بدست آمده است:

برای OT1CTLها مقدار $Kin(t, p) = 1.3767 \times 10^9 \text{ Cell/Kg}$ و برای آنتی‌بادی‌ها مقدار $Kin(t, q) = 1.355 \times 10^{109} \text{ mg/ml}$ در نظر گرفته شده است.

فصل ۸ - نتیجه‌گیری و پیشنهادات

در این رساله با استفاده از مفاهیم یادگیری تقویتی کنترل تحمل‌پذیر عیب برای سیستم‌های دینامیکی مورد بحث قرار گرفت. در حالت کلی سه روش کلی در این رساله مطرح گردید که مورد اول، کنترل تحمل‌پذیر عیب با استفاده از یادگیری تقویتی و الگوریتم گام خودکار بود. در این روش، ابتدا، رفتار سیستم حلقه بسته با استفاده از یک شبکه عصبی پیشرو که با قانون پس انتشار خطا آموزش می‌یافت، شناسایی گردید. در زمان شناسایی سیستم به کمک شبکه عصبی فرض بر این بود که سیستم در حالت نرمال در حال کار کردن می‌باشد و شبکه عصبی رفتار سیستم نرمال را بدون رخداد عیب یاد می‌گرفت. پس از یادگیری شبکه و شناسایی مدل سیستم کنترلی، خروجی شبکه عصبی با خروجی واقعی سیستم مقایسه می‌گردید و مانده استخراج می‌گردید در صورتی که دامنه این مانده از مقداری بیشتر بود آنگاه سیستم کنترلی اعلام رخداد عیب می‌کرد و سپس پارامترهای کنترل کننده در جهت کنترل کردن عیب تغییر می‌کرد. این الگوریتم بر روی مدل تاخیر دار پالمبو شبیه‌سازی گردید و نشان داده شد که در تشخیص عیب دقت بالایی دارد و تا حد زیادی قابلیت کنترل تحمل‌پذیر عیب را نیز دارا می‌باشد. سپس کنترل تحمل‌پذیر عیب با ترکیب یادگیری تقویتی و کنترل مد لغزشی انجام گردید. در این قسمت نیز ابتدا رفتار سیستم نرمال در عدم حضور عیب با استفاده از یک شبکه عصبی RBF شناسایی گردید. سپس در صورت رخداد عیب با مقایسه خروجی شبکه عصبی و سیستم واقعی، مانده محاسبه می‌گردید. علاوه بر این سیگنال دیگری که تفاوت سیستم واقعی و مقدار مطلوب بود به کنترلر مد لغزشی داده می‌شد. در این بین الگوریتم مسیرهای شایستگی با دریافت حالت‌های سیستم واقعی و تولید سیگنال کنترلی مربوطه به کنترلر مد لغزشی در جهت تحمل عیب کمک می‌کرد. شایان ذکر است که در حالت استفاده از کنترل مد لغزشی به تنهایی اولاً سیستم کنترلی قادر به تحمل عیب به خوبی نبود و همچنین در پاسخ سیستم چترینگ زیادی دیده می‌شد اما با ترکیب کنترل مد لغزشی با یادگیری تقویتی و مسیرهای شایستگی چترینگ تا حد

زیادی برطرف می‌گردید و سیستم رفتار مناسب‌تر در رخداد عیب از خود نشان می‌داد. این روش که در فصل شش مورد بحث قرار گرفت بر روی دو سیستم دیابت شبیه‌سازی گردید و رفتار مناسب آن در کنترل تحمل‌پذیر عیب مشاهده گردید. در قسمت آخر این رساله نیز عملکرد کنترل تحمل‌پذیر خطای الگوریتم یادگیری تقویتی به تنهایی مورد بحث قرار گرفت در این فصل ابتدا دامنه و نوع عیب با استفاده از یک روش جدید مبتنی بر شبکه‌های عصبی پیشرو مشخص می‌گردید و علاوه بر دامنه و نوع عیب، سیگنال تفاوت ورودی مطلوب و خروجی واقعی سیستم و حالت‌های سیستم به کنترلر RL داده می‌شد و کنترلر بر اساس این داده‌ها قابلیت خود در کنترل تحمل‌پذیر خطا را نشان می‌داد. در این بخش نیز برای آنکه مقایسه مناسبی مابین این روش و روش‌های قبلی صورت گیرد روش مذکور با روش ترکیبی کنترل مد لغزشی و مسیر شایستگی که در فصل قبل آن مورد بحث قرار گرفته بود مقایسه گردید. برای آنکه مقایسه تا حد زیادتری عادلانه باشد دامنه و نوع عیب که توسط شبکه عصبی محاسبه می‌گردید به قسمت RL داده شد. در نهایت پس از شبیه‌سازی بر روی مدل دیابت مشاهده گردید که کنترلر پیشنهادی فصل هفت در عین حالیکه بسیار ساده‌تر نسبت به کنترلرهای قبلی می‌باشد نشان داده شد که در حدود $5 \frac{mg}{dl}$ به مقدار مطلوب نزدیک‌تر می‌باشد و با سرعت بیشتری نیز به همگرایی رسیده است. در مورد انسولین نیز این ساختار پیشنهادی به صورت میانگین PM ۴ به مقدار مطلوب نزدیک‌تر می‌باشد.

در میان سه روش پیشنهادی در این رساله هر سه روش قابلیت کنترل تحمل‌پذیر عیب برای عیب‌های ناگهانی را دارا می‌باشند اما در برابر عیب‌های جمع شونده نرم فقط روش پنجره زمانی قابلیت کنترل تحمل‌پذیر عیب را دارا می‌باشد. در روش پیشنهادی ترکیب مسیرهای شایستگی و کنترل مد لغزشی سرعت و دقت کنترل تحمل‌پذیر عیب برای عیب‌های ناگهانی نسبت به روش‌های پیشین بهبود داده شد و علاوه بر این از آنجا که در یادگیری تقویتی ماهیت ذاتی کنترل بهینه وجود دارد با کم شدن میزان دارو در عین کنترل بقیه موارد بیماری‌ها اثرات سوء دارو نیز کم می‌شود، که این مساله مهمی در علم پزشکی می‌باشد.

به منظور ادامه تحقیق در آینده موارد زیر پیشنهاد می‌گردد:

- پیشنهاد می‌گردد که کنترل تحمل‌پذیر عیب با استفاده از روش‌های یادگیری تقویتی ایمن^۱ انجام شود بدین ترتیب که چندین کنترل‌کننده بر اساس تابع لیاپانوف طراحی گردد و یادگیری تقویتی مسئولیت سوییچ کردن بین این کنترل‌کننده‌ها را داشته باشد تا با دقت بیشتری به مقدار مطلوب برسد.
- یادگیری تقویتی که در این رساله مورد استفاده قرار گرفت یادگیری تقویتی گسسته در زمان با در نظر گرفتن حالت‌ها و عمل‌های گسسته بود در صورتی که تعداد بازه‌های حالت‌ها و عمل‌ها از یک حدی بیشتر شود مساله دچار نفرین ابعاد می‌شود و این روش یادگیری تقویتی پاسخگو نمی‌باشد پیشنهاد می‌شود در کارهای آینده از الگوریتم‌های یادگیری تقویتی پیوسته که برای حالت‌های پیوسته تعریف می‌شود استفاده گردد. بدین منظور می‌بایست مفاهیم یادگیری تقویتی با شبکه‌های عصبی مصنوعی ترکیب گردد. تا قابلیت کار کردن در فضای پیوسته را دارا باشد. همچنین پیشنهاد می‌شود برای فضاهای بسیار بزرگ از یادگیری تقویتی عمیق استفاده گردد.
- مورد بعد اینکه در الگوریتم‌های یادگیری تقویتی می‌توان از سیاست‌های جدید همانند سیاست فراابتکارانه بهره برد. یا روش‌های دیگر یادگیری تقویتی همانند یادگیری تقویتی فئودال مورد استفاده قرار گیرد و با نتایج این رساله مقایسه گردد.

¹ Safe Reinforcement learning

- [١] R. Isermann, *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer Science & Business Media, 2006.
- [٢] H. Alwi, C. Edwards, and C. P. Tan, *Fault detection and fault-tolerant control using sliding modes*. Springer Science & Business Media, 2011.
- [٣] A. Milhim, Y. Zhang, and C.-A. Rabbath, "Gain scheduling based PID controller for fault tolerant control of quad-rotor UAV," in *AIAA infotech@ aerospace 2010*, 2010, p. 3530.
- [٤] N. Jin, S. Hu, C. Gan, and Z. Ling, "Finite states model predictive control for fault-tolerant operation of a three-phase bidirectional AC/DC converter under unbalanced grid voltages," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 1, pp. 819-829, 2017.
- [٥] P. Mhaskar, "Robust model predictive control design for fault-tolerant control of process systems," *Industrial & engineering chemistry research*, vol. 45, no. 25, pp. 8565-8574, 2006.
- [٦] X. Hu, H. R. Karimi, L. Wu, and Y. Guo, "Model predictive control-based non-linear fault tolerant control for air-breathing hypersonic vehicles," *IET Control Theory & Applications*, vol. 8, no. 13, pp. 1147-1153, 2014.
- [٧] M. Staroswiecki, "Fault tolerant control: the pseudo-inverse method revisited," *Ifac proceedings volumes*, vol. 38, no. 1, pp. 418.٢٠٠٥ ,٤٢٣-
- [٨] H. Guo, J. Xu, and Y.-H. Chen, "Robust control of fault-tolerant permanent-magnet synchronous motor for aerospace application with guaranteed fault switch process," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 12, pp. 7309-7321.٢٠١٥ ,
- [٩] Y. Shen, L. Liu, and E. H. Dowell, "Adaptive fault-tolerant robust control for a linear system with adaptive fault identification," *IET Control Theory & Applications*, vol. 7, no. 2, pp. 246-252, 2013.
- [١٠] K. Zhou and Z. Ren, "A new controller architecture for high performance, robust, and fault-tolerant control," *IEEE Transactions on automatic control*, vol. 46, no. 10, pp. 1613-1618, 2001.
- [١١] Z. Qu, C. M. Ihlefeld, Y. Jin, and A. Saengdeejing, "Robust fault-tolerant self-recovering control of nonlinear uncertain systems," *Automatica*, vol. 39, no. 10, pp. 1763-1771, 2003.
- [١٢] D. Ye and G.-H. Yang, "Adaptive fault-tolerant tracking control against actuator faults with application to flight control," *IEEE Transactions on control systems technology*, vol. 14, no. 6, pp. 1088-1096, 2006.
- [١٣] Z. Zuo, J. Zhang, and Y. Wang, "Adaptive fault-tolerant tracking control for linear and Lipschitz nonlinear multi-agent systems," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3923-3931.٢٠١٤ ,
- [١٤] B. Xiao, Q. Hu, and Y. Zhang, "Adaptive sliding mode fault tolerant attitude tracking control for flexible spacecraft under actuator saturation," *IEEE Transactions on Control Systems Technology*, vol. 20, no. 6, pp. 1605-1612, 2011.
- [١٥] H. Li ,H. Gao, P. Shi, and X. Zhao, "Fault-tolerant control of Markovian jump stochastic systems via the augmented sliding mode observer approach," *Automatica*, vol. 50, no. 7, pp. 1825-1834, 2014.
- [١٦] M. Van, S. S. Ge, and H. Ren, "Finite time fault tolerant control for robot manipulators using time delay estimation and continuous nonsingular fast terminal sliding mode control," *IEEE transactions on cybernetics*, vol. 47, no. 7, pp. 1681-1693, 2016.

- [١٧] A. Vargas-Martinez and L. Garza-Castañón, "Combining artificial intelligence and advanced techniques in fault-tolerant control," *Journal of applied research and technology*, vol. 9, no. 2, pp. 202-226, 2011.
- [١٨] H. Wang and Y. Wang, "Neural-network-based fault-tolerant control of unknown nonlinear systems," *IEE Proceedings-Control Theory and Applications*, vol. 146, no. 5, pp. 389-398, 1999.
- [١٩] X.-J. Li and G.-H. Yang, "Neural-network-based adaptive decentralized fault-tolerant control for a class of interconnected nonlinear systems," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 1, pp. 144-155, 2016.
- [٢٠] H. Jiang, H. Zhang, Y. Liu, and J. Han, "Neural-network-based control scheme for a class of nonlinear systems with actuator faults via data-driven reinforcement learning method," *Neurocomputing*, vol. 239, pp. 1-8, 2017.
- [٢١] L.-Y. Hao, J. H. Park, and D. Ye, "Fuzzy logic systems-based integral sliding mode fault-tolerant control for a class of uncertain non-linear systems," *IET Control Theory & Applications*, vol. 10, no. 3, pp. 300-3٠١٦, ١١.
- [٢٢] C. Lopez-Toribio, R. J. Patton, and S. Daley, "Takagi-Sugeno fuzzy fault-tolerant control of an induction motor," *Neural Computing & Applications*, vol. 9, no. 1, pp. 19-28, 2000.
- [٢٣] F.-J. Lin, Y.-C. Hung, J.-C. Hwang, and M.-T. Tsai, "Fault-tolerant control of a six-phase motor drive system using a Takagi-Sugeno-Kang type fuzzy neural network with asymmetric membership function," *IEEE Transactions on Power Electronics*, vol. 28, no. 7, pp. 3557-3572, 2012.
- [٢٤] L. Liu, Z. Wang, and H. Zhang, "Adaptive fault-tolerant tracking control for MIMO discrete-time systems via reinforcement learning algorithm with less learning parameters," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 1, pp. 299-313, 2016.
- [٢٥] Z. Wang, L. Liu, H. Zhang, and G. Xiao, "Fault-tolerant controller design for a class of nonlinear MIMO discrete-time systems via online reinforcement learning algorithm," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 5, pp. 611-622, 2015.
- [٢٦] F. Shi, "Observer based active fault tolerant control of descriptor systems," University of Hull Kingston upon Hull, UK, 2013 .
- [٢٧] J. Skach, I. Punčochář, and F. L. Lewis, "Temporal-difference Q-learning in active fault diagnosis," in *2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol)*, 2016: IEEE, pp. 287-292 .
- [٢٨] Y. Y. Hsu and C. C. Yu, "A self-learning fault diagnosis system based on reinforcement learning," *Industrial & engineering chemistry research*, vol. 31, no. 8, pp. 1937 ,١٩٤٦-١٩٩٢.
- [٢٩] A. Förster and A. L. Murphy, "Froms: A failure tolerant and mobility enabled multicast routing paradigm with reinforcement learning for WSNs," *Ad Hoc Networks*, vol. 9, no. 5, pp. 940-965, 2011.
- [٣٠] R. J. Patton, F. J. Uppal, S. Simani, and B. Polle, "A Monte Carlo analysis and design for FDI of a satellite attitude control system," *IFAC Proceedings Volumes*, vol. 39, no. 13, pp. 1318-1323, 2006.
- [٣١] R. S. Istepanian, N. Y. Philip, and M. G. Martini, "Medical QoS provision based on reinforcement learning in ultrasound streaming over 3.5 G wireless systems," *IEEE Journal on Selected areas in Communications*, vol. 27, no. 4, pp. 566-574, 2009.
- [٣٢] C. C. Johnson, "Parallel Fault Tolerant Multi-Agent Reinforcement Learning," 2011.
- [٣٣] J. Škach and I. Punčochář, "Input design for fault detection using extended kalman filter and reinforcement learning," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7302-7307, 2017.
- [٣٤] F. Farivar and M. N. Ahmadi, "Continuous reinforcement learning to robust fault tolerant control for a class of unknown nonlinear systems," *Applied Soft Computing*, vol. 37, pp. 702-714, 2015.

- [٣٥] K.-Z. Han, J. Feng, and X. Cui, "Fault-tolerant optimised tracking control for unknown discrete-time linear systems using a combined reinforcement learning and residual compensation methodology," *International Journal of Systems Science*, vol. 48, no. 13, pp. 2811-2825, 2017.
- [٣٦] D. Zhang and Z. Gao, "Reinforcement learning-based fault-tolerant control with application to flux cored wire system," *Measurement and Control*, vol. 51, no. 7-8, pp. 349-359, 2018.
- [٣٧] H. H. Afshari, D. Al-Ani, and S. Habibi, "Fault Prognosis of Roller Bearings Using the Adaptive Auto-Step Reinforcement Learning Technique," in *ASME 2014 Dynamic Systems and Control Conference*, 2014: American Society of Mechanical Engineers Digital Collection .
- [٣٨] S. Shamshirband, A. Patel, N. B. Anuar, M. L. M. Kiah, and A. Abraham, "Cooperative game theoretic approach using fuzzy Q-learning for detecting and preventing intrusions in wireless sensor networks," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 228-241, 2014.
- [٣٩] K. Basterretxea, M. V. Martínez, I. Del Campo, and J. Echanobe, "A fault tolerant single-chip intelligent agent with feature extraction capability," *Applied Soft Computing*, vol. 22, pp. 358-371, 2014.
- [٤٠] S. R. Ahmadzadeh, P. Kormushev, and D. G. Caldwell, "Multi-objective reinforcement learning for AUV thruster failure recovery," in *2014 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, 2014: IEEE, pp. 1-8 .
- [٤١] C. Hua, S. X. Ding, and Y. A. Shardt, "A New Method for Fault Tolerant Control through Q-Learning," *IFAC-PapersOnLine*, vol. 51, no. 24, pp. 38-45, 2018.
- [٤٢] Y. Zhang and J. Jiang, "Bibliographical review on reconfigurable fault-tolerant control systems," *Annual reviews in control*, vol. 32, no. 2, pp. 229-252, 2008.
- [٤٣] R. J. Patton, P. M. Frank, and R. N. Clarke, *Fault diagnosis in dynamic systems: theory and application*. Prentice-Hall, Inc., 1989.
- [٤٤] Y. Zhang and J. Jiang, "Bibliographical review on reconfigurable fault-tolerant control systems," *IFAC Proceedings Volumes*, vol. 36, no. 5, pp. 257-268, 2003.
- [٤٥] J. Chen and R. J. Patton, *Robust model-based fault diagnosis for dynamic systems*. Springer Science & Business Media, 2012.
- [٤٦] L. Garza-Castañón and A. Vargas-Martínez, "Artificial Intelligence Methods in Fault Tolerant Control, Chapter of Book Automation and Control-Theory and Practice, Ed," *TECH, Austria*, 2010.
- [٤٧] C. D. Bocaniala and V. Palade, "Computational intelligence methodologies in fault diagnosis: review and state of the art," in *Computational intelligence in fault diagnosis*: Springer, 2006, pp. 1-36.
- [٤٨] J. Korbicz, J. M. Koscielny, Z. Kowalczyk, and W. Cholewa, *Fault diagnosis: models, artificial intelligence, applications*. Springer Science & Business Media, 2012.
- [٤٩] M. Kowal and J. Korbicz, "Robust fault detection using neuro-fuzzy networks," *IFAC Proceedings Volumes*, vol. 38, no. 1, pp. 185-190, 2005.
- [٥٠] R. J. Patton, J. Chen, and J. Chen, "A study on neuro-fuzzy systems for fault diagnosis," *International Journal of Systems Science*, vol. 31, no. 11, pp. 1441-1448, 2000.
- [٥١] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction," 2011.
- [٥٢] M. Littman and A. Moore, "Reinforcement Learning: A Survey, Journal of Artificial Intelligence Research 4," ed: syf, 1996.
- [٥٣] R. S. a. A. Barto., "Reinforcement learning: An introduction," *The MIT Press Cambridge, Massachusetts London, England*, 2011.
- [٥٤] R. M. R. Bianchi, "Case-Based Multi agent Reinforcement Learning: Cases as Heuristics for Selection of Actions," *ECAI*, vol. 215, pp. 355-360, 2010.
- [٥٥] C. J. C. H. Watkins, "Learning from delayed rewards," 1989.
- [٥٦] N. C. Van Der Ven *et al.*, "The confidence in diabetes self-care scale: psychometric properties of a new measure of diabetes-specific self-efficacy in Dutch and US patients with type 1 diabetes," *Diabetes care*, vol. 26, no. 3, pp. 713-718, 2003.

- [٥٧] A. A. Sharief and A. Sheta, "Developing a mathematical model to detect diabetes using multigene genetic programming," *IJARAI) International Journal of Advanced Research in Artificial Intelligence*, vol. 3, no. 10, 2014.
- [٥٨] A. Makroglou, I. Karaouostas, J. Li, and Y. Kuang, "Delay differential equation models in diabetes modeling," *Theoretical Biology and Medical Modelling*, 2009.
- [٥٩] P. Palumbo, S. Panunzi, and A. De Gaetano, "Qualitative behavior of a family of delay-differential models of the glucose-insulin system," *Discrete and Continuous Dynamical Systems Series B*, vol. 7, no. 2, p. 399, 2007.
- [٦٠] I. Pagkalos, P. Herrero, C. Toumazou, and P. Georgiou, "Bio-Inspired Glucose Control in Diabetes Based on an Analogue Implementation of a β -Cell Model," *IEEE transactions on biomedical circuits and systems*, vol. 8, no. 2, pp. 186-195, 2014.
- [٦١] Z. Tashakorizade, N. Naghavi, and S. H. Sani, "Glucose regulation in type 1 diabetes mellitus with model reference adaptive control and modified smith predictor," *Iranian Journal of Biomedical Engineering*, vol. 8, pp. 159-171, 2014.
- [٦٢] M. S. Akhyani and H. Khaloozadeh, "Optimal glucose-insulin regulatory system in type1 diabetic patients based on the nonlinear time delay models," *Journal of Control*, vol. 8, no. 4, pp. 31-41, 2015.
- [٦٣] P. Latafat, P. Palumbo, P. Pepe, L. Kovác, S. Panunzi, and A. De Gaetano, "An LMI-based controller for the glucose-insulin system," in *2015 European Control Conference (ECC)*, 2015: IEEE, pp. 7-12 .
- [٦٤] N. Magdelaine *et al.*, "A long-term model of the glucose–insulin dynamics of type 1 diabetes," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 6, pp. 1546-1552, 2015.
- [٦٥] F. Piccinini, C. Dalla Man, A. Vella, and C. Cobelli, "A model for the estimation of hepatic insulin extraction after a meal," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 9, pp. 1925-1932, 2015.
- [٦٦] M. K. Bothe *et al.*, "The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas," *Expert review of medical devices*, vol. 10, no. 5, pp. 6٠١٣, ٦٧٣-٦١, 2013.
- [٦٧] P. Grant, "A new approach to diabetic control: fuzzy logic and insulin pump technology," *Medical engineering & physics*, vol. 29, no. 7, pp. 824-827, 2007.
- [٦٨] D. U. Campos-Delgado, M. Hernández-Ordoñez, R. Femat, and A. Gordillo-Moscoso, "Fuzzy-based controller for glucose regulation in type-1 diabetic patients by subcutaneous route," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2201-2210, 2006.
- [٦٩] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," *Computational and structural biotechnology journal*, vol. 15, pp. 104-116, 2017.
- [٧٠] I. Hochberg, G. Feraru, M. Kozdoba, S. Mannor, M. Tennenholtz, and E. Yom-Tov, "A reinforcement learning system to encourage physical activity in diabetes patients," *arXiv preprint arXiv:1605.04070*, 2016.
- [٧١] S. Yasini, M. Naghibi-Sistani, and A. Karimpour, "Agent-based simulation for blood glucose control in diabetic patients," *International Journal of Applied Science, Engineering and Technology*, vol. 5, no. 1, pp. 40-49, 2009.
- [٧٢] W.-H. Weng, M. Gao, Z. He, S. Yan, and P. Szolovits, "Representation and reinforcement learning for personalized glycemic control in septic patients," *arXiv preprint arXiv:1712.00654*, 2017.
- [٧٣] I. Contreras and J. Vehi, "Artificial intelligence for diabetes management and decision support: literature review," *Journal of medical Internet research*, vol. 20, no. 5, p. e10775, 2018.
- [٧٤] A. Roy and R. S. Parker, "Dynamic modeling of exercise effects on plasma glucose and insulin levels," ed: SAGE Publications, 2007.

- [٧٥] P. Magni and R. Bellazzi, "A stochastic model to assess the variability of blood glucose time series in diabetic patients self-monitoring," *IEEE Transactions on biomedical engineering*, vol. 53, no. 6, pp. 977-985, 2006.
- [٧٦] A. Makroglou, J. Li, and Y. Kuang, "Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview," *Applied numerical mathematics*, vol. 56, no. 3-4, pp. 559-573, 2006.
- [٧٧] Y. C. Kueh, T. Morris, E. Borkoles, and H. Shee, "Modelling of diabetes knowledge, attitudes, self-management, and quality of life: a cross-sectional study with an Australian sample," *Health and quality of life outcomes*, vol. 13, no. 1, p. 129, 2015.
- [٧٨] F. Nani and M. Jin, "Mathematical modeling and simulations of the pathophysiology of Type-2 Diabetes Mellitus," in *2015 8th International Conference on Biomedical Engineering and Informatics (BMEI)*, 2015: IEEE, pp. 296-300 .
- [٧٩] J. R. Moore and F. Adler, "Mathematical modeling of type 1 diabetes in the NOD mouse: separating incidence and age of onset," *arXiv preprint arXiv:1412.6566*, 2014.
- [٨٠] A. Mahata, S. P. Mondal, S. Alam, and B. Roy, "Mathematical model of glucose-insulin regulatory system on diabetes mellitus in fuzzy and crisp environment," *Ecological Genetics and Genomics*, vol. 2, pp. 25-34, 2017.
- [٨١] A. Onvlee, H. Blauw, N. Middelhuis, and H. Zwart, "In silico modeling of patients with type 1 diabetes mellitus," MS thesis, Dep. Tech. Med., Univ. Twente, Enschede, Netherlands, 2016 .
- [٨٢] D. Sandhya danKumar, "Mathematical Model for GlucoseInsulinRegulatorySystem of Diabetes Mellitus," *Advances in applied mathematical biosciences*, vol. 2, no. 1 ,pp. 39-46, 2011.
- [٨٣] A. De Gaetano *et al.*, "Mathematical models of diabetes progression," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 295, no. 6, pp. E1462-E1479, 2008.
- [٨٤] S. A. Gandhi and J. Kampp, "Skin cancer epidemiology, detection, and management," *Medical Clinics*, vol. 99, no. 6, pp. 1323-1335, 2015.
- [٨٥] R. A. Schwartz, *Skin cancer: recognition and management*. John Wiley & Sons, 2008.
- [٨٦] S. Eikenberry, C. Thalhauser, and Y. Kuang, "Tumor-immune interaction, surgical treatment, and cancer recurrence in a mathematical model of melanoma," *PLoS computational biology*, vol. 5, no. 4, p. e1000362, 2009.
- [٨٧] Y. Kogan, Z. Agur, and M. Elishmereni, "A MATHEMATICAL MODEL FOR THE IMMUNOTHERAPEUTIC CONTROL OF THE TH1/TH2 IMBALANCE IN MELANOMA," *Discrete & Continuous Dynamical Systems-Series B*, vol. 18, no. 4, 2013.
- [٨٨] L. DePillis, A. Gallegos, and A. Radunskeya, "A model of dendritic cell therapy for melanoma," *Frontiers in oncology*, vol. 3, 2013.
- [٨٩] X. Sun, J. Bao, and Y. Shao, "Mathematical modeling of therapy-induced cancer drug resistance: connecting cancer mechanisms to population survival rates," *Scientific reports*, vol. 6, p. 22498, 2016.
- [٩٠] W. Dzwiniel, A. Klusek, and O. V. Vasilyev, "Supermodeling in simulation of melanoma progression," *Procedia Computer Science*, vol. 80, pp. 999-1010, 2016.
- [٩١] A. I. Mendes, C. Nogueira, J. Pereira, and R. Fonseca-Pinto, "On the geometric modulation of skin lesion growth: a mathematical model for melanoma," *Research on Biomedical Engineering*, vol. 32, no. 1, pp. 44-54, 2016.
- [٩٢] M. Pennisi, "A mathematical model of immune-system-melanoma competition," *Computational and mathematical methods in medicine*, vol. 2012, 2012.
- [٩٣] H. B. Demuth, M. H. Beale, O. De Jess, and M. T. Hagan, *Neural network design*. Martin Hagan, 2014.
- [٩٤] A. R. Mahmood, R. S. Sutton, T. Degris, and P. M. Pilarski, "Tuning-free step-size adaptation," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012: IEEE, pp. 2121-21 .٢٤

- [٩٥] R. S. Sutton, "Adapting bias by gradient descent: An incremental version of delta-bar-delta," in *AAAI*, 1992, pp. 171-176 .
- [٩٦] K. Parand, S. Abbasbandy, S. Kazem, and A. Rezaei, "Comparison between two common collocation approaches based on radial basis functions for the case of heat transfer equations arising in porous medium," *Communications in Nonlinear Science Numerical Simulation* vol. 16, no. 3, pp. 1396-1407, 2011.
- [٩٧] T. Carmenate *et al.*, "Blocking IL-2 signal in vivo with an IL-2 antagonist reduces tumor growth through the control of regulatory T cells," *The Journal of Immunology*, vol. 200, no. 10, pp. 3475-3484, 2018.
- [٩٨] N. A. Alias, "Linear Quadratic Regulator (LQR) controller design for Inverted Pendulum," Universiti Tun Hussein Onn Malaysia, 2013 .

Abstract

This thesis investigates the fault tolerant control from different aspects, by using the Reinforcement Learning methods. First, the robustness of the Reinforcement Learning is studied. It is demonstrated that the represented approach has the robustness feature. Then, through identifying the fault-free (normal) system by using the Radial Basis Function (RBF), the residual is calculated. The main innovation of this thesis is combining the Sliding Mode control concept and the Eligibility Traces algorithm, one of the most applicable algorithms in the Reinforcement Learning framework, for the fault tolerant control. The proposed structure is compared with the previous one, a combination of the Artificial Neural Networks and the Sliding Mode control. It is shown that the performance of the proposed structure is superior than the previous one. The second innovation of the thesis is using the Time Window to calculate the amplitude and determining the type of the fault. The resulting data and states of the system are used to implement the fault tolerant control by using the Eligibility Traces algorithm. Finally, it is revealed that the proposed structure has a better performance in addition to the simplicity of the tolerant control. The basic methods of the Temporal Difference learning, including the Q-learning algorithm and SARSA are compared to demonstrate the fault tolerant control behavior of the proposed structure. In The last innovation, the Feedforward neural network is used to identify the normal and faulty systems, based on the residual calculation. Based on the calculated residual, continuous Reinforcement Learning in combination with the Auto-Step algorithm is employed for the fault tolerant control purpose. Proposed methods are applied to diabetic and melanoma patient systems. There is no need of the environment model in RL and it is one of the most important features of the Reinforcement Learning algorithm in the system control. Since there is no access to the real model (patient), the mathematical models of diabetic and melanoma patients are utilized. To treat the melanoma in the presence of the fault, the half-life of the drug is considered, as well. This is an important consideration for problems which contribute with determining the optimal dosage and studied in a few researches.

Keywords:

Fault Tolerant Control, Reinforcement Learning, Eligibility Traces Algorithm, Artificial Neural Networks, Q-learning Algorithm



Shahrood University of Technology

Faculty of Electrical and Robotics Engineering

M.Sc. Thesis in Control Engineering

**Fault detection and Fault tolerant control for
dynamic systems using reinforcement learning**

By

Amin Noori

Supervisor:

Dr. M. A.Sadrnia

Advisor:

Dr. M. B. Naghibi Sistani

February 2020