

بررسی مدیریت کلان داده در تداخل دارو با دارو براساس رویکرد یادگیری عمیق مدرن برای پزشکی و مراقبت های بهداشتی هوشمند

مهدی میرمعصومی^{۱*}

^۱ کارشناسی ارشد، گروه حسابداری، موسسه آموزش عالی شمیم دانش نوین، اردبیل، ایران

چکیده

تشخیص و طبقه‌بندی تداخلات دارو با دارو (DDI) از داده‌های موجود از اهمیت بالایی برخوردار است زیرا گزارش‌ها نشان می‌دهد که DDI یکی از دلایل اصلی شرایط اکتسابی در بیمارستان است. همچنین برای مراقبت‌های بهداشتی هوشمند برای جلوگیری از تداخلات دارویی نامطلوب، داشتن دانش به روز در مورد DDI ضروری است. این دانش را می‌توان با استفاده از تکنیک‌های پردازش متن در ادبیات پزشکی در قالب «کلان داده» استخراج کرد. ادبیات پزشکی در حجم بسیار زیادی منتشر می‌شود، و جمع‌آوری تمام گزارش‌های بررسی شده از این کلان داده برای متخصصان مراقبت‌های بهداشتی غیرممکن است. برای جلوگیری از این روش زمان‌بر، تکنیک‌های استخراج اطلاعات (IE) و استخراج رابطه (RE) که به طور عمیق در پردازش زبان طبیعی (NLP) مورد مطالعه قرار گرفته‌اند. پیشرفته‌ترین پیاده‌سازی‌های فعلی برای استخراج DDIها از متون زیست پزشکی، از ماشین‌های بردار پشتیبان (SVM) یا سایر روش‌های یادگیری ماشینی استفاده می‌کنند که بر روی ویژگی‌های تعریف شده دستی کار می‌کنند. تکنیک‌های مدرن یادگیری عمیق نیز برای استخراج خودکار DDIها از متون علمی به کار گرفته شده‌اند و ثابت کرده‌اند که برای پیشرفت وظایف استخراج DDI بسیار امیدوارکننده هستند. ما یک روش مبتنی بر شبکه عصبی عمیق (SEV-DDI: شدت تداخل دارو) با چند لایه یکپارچه‌تر برای دستیابی به دقت و دقت بالاتر پیشنهاد کردیم. توانایی تعیین شدت DDI برای سیستم‌های پشتیبانی تصمیم‌گیری بالینی دقیق‌تر و آگاهانه‌تر بسیار مفید خواهد بود.

واژه‌های کلیدی: تداخل دارو با دارو، استخراج اطلاعات، یادگیری عمیق، پزشکی، مراقبت‌های بهداشتی هوشمند

۱. مقدمه

در چند دهه اخیر، تحقیقات زیست پزشکی و مراقبت های بهداشتی هوشمند از نظر حضور در ادبیات، اکتشافات جدید و رویکردهای محاسباتی شاهد رشد سریعی بوده است که به همین دلیل حجم عظیمی از داده های تجربی تولید و منتشر شده است. (به عنوان مثال، داده های بزرگ) برای تأیید و توصیف این نوآوری ها [۱]. حجم داده ها به قدری زیاد است که تجزیه و تحلیل و مطالعه مقالات مرتبط با رشته مورد نظر خود برای انسان غیرممکن است و حتی استخراج ساده مقالات منتشر شده مرتبط با حوزه تقریباً غیرممکن شده است. به این ترتیب، برای استفاده از اطلاعات نشریات و مقالات علمی، افزایش استفاده از اتوماسیون به دلیل تنوع و رشد انفجاری در ادبیات مراقبت های بهداشتی و همچنین در صنعت داروسازی، نمایانگر واضح این رشد است. به همین دلیل است که اکنون توجه به پیاده سازی و تکامل ابزارهای کشف مبتنی بر دانش منحرف شده و متمرکز شده است: حجم عظیم تحقیقات زیست پزشکی در ادبیات به اتوماسیون برای استخراج، بازنمایی، تفسیر و نگهداری این اطلاعات به شیوه ای تصفیه شده نیاز دارد.

به طور کلی، سازمان های مختلفی در بخش مراقبت های بهداشتی کار می کنند، که نمونه هایی از آن ها عبارتند از سازمان بهداشت جهانی (WHO)، کمیته محیط زیست و سلامت اروپا (EEHC)، و سازمان بهداشت پان آمریکا (PAHO). انگیزه این سازمان ها تضمین ایمنی بیمار است که بالاترین اولویت تلقی می شود و برای دستیابی به این هدف، این سازمان ها فعالانه در اتخاذ تدابیر قانونی برای کنترل اثرات نامطلوب که از طریق استفاده قابل اجتناب است، مشارکت دارند.

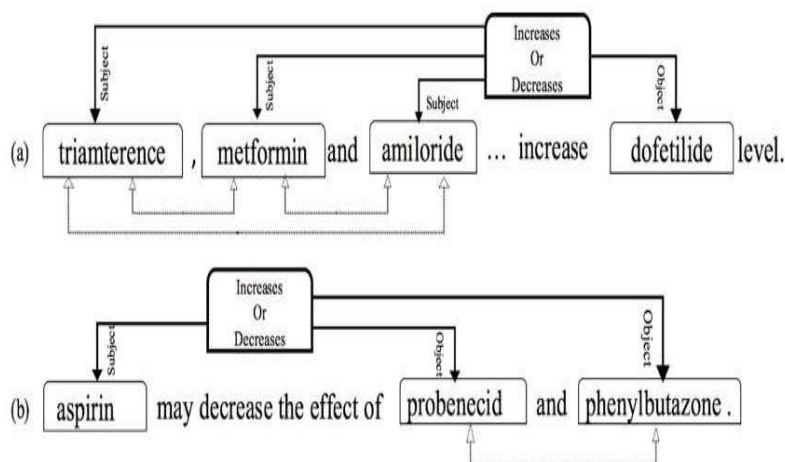
مؤسسه پزشکی (IOM) در ایالات متحده تعیین کرد که سالانه میلیون ها بیمار در بیمارستان ها به دلیل مشکلات مربوط به کیفیت مراقبت از جمله اشتباهات دارویی، عدم تمیزی، عفونت بیمارستانی، زنان و زایمان جان خود را از دست می دهند. و DDI هایی که باعث واکنش های جانبی دارویی (ADRs) می شوند [۴]، [۸]. اگرچه سیستم های به خوبی تثبیت شده برای حفظ ایمنی بیمار وجود دارد، اما زمینه تداخلات دارویی نیاز به تحقیقات بیشتر دارد [۹] زیرا گزارش های اخیر نشان می دهد که واکنش های دارویی یکی از دلایل اصلی شرایط اکتسابی در بیمارستان و بستری مجدد است [۱۰]، [۱۴]. مراقبت دارویی، که توسط WHO آغاز شد، توجه بیشتری را به خود جلب می کند، با ادبیاتی که نشان می دهد تقریباً یک دهم از واکنش های نامطلوب مشاهده شده در افراد ICU به دلیل DDI هایی است که باعث ADR می شوند [۱۵]، [۱۶]. سیستم های ارزیابی و اتوماسیون مختلفی برای تشخیص ADR در داروهای مختلف استفاده می شوند، اما در بیشتر موارد، گزارش های ADR کمی بعد توسط متخصصان مراقبت های بهداشتی بررسی و آشکار می شوند.

انگیزه این پژوهش ایجاد استراتژی های جدید در سیستم های از قبل فعال برای تشخیص DDI از ادبیات پزشکی با در نظر گرفتن «داده های بزرگ» است. طبق تعریف، یک دارو ممکن است اثر داروی دیگری را به دلیل فرمول شیمیایی متقابل آن کاهش یا افزایش دهد، پدیده ای که به عنوان DDI شناخته می شود [۱۷]، [۱۹]. اطلاعات مربوط به DDI ها برای محدود کردن شکست های درمان های درمانی و جلوگیری از سایر ADR های قوی بسیار مهم و مرتبط است. اگرچه DDI ها می توانند بسیار خطرناک باشند، این در همه موارد صادق نیست و این فعل و انفعالات می توانند در موارد خاصی که اثر هم افزایی مطلوبی را ایجاد می کنند مؤثر باشند [۲۰]. بنابراین، ما علاقه مندیم که شدت یک DDI را با پردازش همان متن مورد استفاده برای تشخیص DDI تعیین کنیم. این جنبه به سیستم های پشتیبانی تصمیم گیری بالینی در تصمیم گیری دقیق تر و آگاهانه تر با اطمینان از ایمنی بیماران کمک می کند. این نه تنها جان انسان ها را نجات می دهد، بلکه منجر به کاهش شدید هزینه های مراقبت های بهداشتی نیز می شود.

DDI تأثیر بسیار زیادی بر ایمنی بیماران دارد و بنابراین باید به طور جدی با آن مقابله شود. رشد سریعی در توسعه فناوری های جدید و سیستم های خبره در حوزه مراقبت های بهداشتی و پزشکی وجود داشته است. با گذشت زمان، نیاز به رویکردهای یکپارچه و مراقبت های بهداشتی مبتنی بر شواهد با سرعت بسیار بالایی در حال افزایش است. از طرف دیگر، اندازه و همچنین ناهمگونی و پیچیدگی داده های تولید شده توسط چندین منبع چالش بزرگی برای رویکردهای محاسباتی است [۲۱]، [۲۴]. درک متن عظیم زیست پزشکی یک چالش برای تحقیقات مراقبت های بهداشتی است [۲۵]. متن بدون ساختار

مهم ترین شکل داده در حوزه های مراقبت های بهداشتی و پزشکی را تشکیل می دهد. اگرچه پردازش این شکل از داده ها بسیار دشوار است، اما شامل جزئیات، نکات ظریف و جزئیات بسیاری است که نمی توان آن ها را در اصطلاحنامه نام گذاری ثبت کرد، اما این اطلاعات می تواند در تصمیم گیری بسیار مفید باشد [۲۶]. مقابله با این اطلاعات بدون ساختار رو به رشد به تکنیک ها و روش های یادگیری ماشینی نیاز دارد تا ساختار و بینش های معنادار استخراج شود [۲۷].

پایگاه های داده تخصصی زیادی وجود دارد (به عنوان مثال، بانک دارو، Micromedex، MEDLINE) که حاوی DDI های شناخته شده است. این جفت های DDI از نظر پوشش محدود هستند زیرا اکتشافات جدید ساخته شده و به عنوان مقالات تحقیقاتی منتشر می شوند. در نتیجه، نیاز نهایی برای استخراج DDI تازه کشف شده در ادبیات علمی وجود دارد [۲۸]، [۲۹]. متن کاوی به عنوان یکی از بهترین تکنیک های موجود شناخته می شود که می تواند برای دستیابی به استخراج رابطه خودکار از متون و مقالات منتشر شده استفاده شود. متن کاوی به طور گسترده برای استخراج رابطه مورد استفاده قرار می گیرد، به عنوان مثال برهمکنش های پروتئین-پروتئین (PPIs) و روابط ژن بیماری، و بنابراین یک پیکره [۳۰] به طور خاص برای وظیفه استخراج DDI شکل می گیرد [۳۱]، [۳۳]. وظیفه استخراج DDI بسیار شبیه به استخراج رابطه است که به طور عمیق در حوزه متن کاوی مطالعه شده است همانطور که در شکل ۱ نشان داده شده است. این شکل کار استخراج رابطه را با نمایش مثال های عملی نشان می دهد. در جمله پردازش شده، هر دارو در ابتدا برچسب گذاری می شود، مانند موضوع و مفعول. پس از به دست آوردن جفت دارو، رابطه بین داروها بررسی می شود و اگر یک جفت دارو در افزایش یا کاهش اثر دیگری نقش داشته باشد، همانطور که در مثال ارائه شده با خطوط توپر نشان داده شده است، به عنوان یک جفت DDI مثبت طبقه بندی می شود. در حالی که جفت های DDI منفی (یعنی داروهایی که روی یکدیگر تأثیری ندارند) با خطوط نقطه چین نشان داده می شوند.



شکل ۱- نمایش ساختار یافته DDI ها

در شکل ۱، ساختار پیچیده DDI ها با دو مثال واقعی کاربردی شامل بندهای DDI های مثبت و منفی که به ترتیب با خطوط توپر و نقطه چین نشان داده شده اند، ارائه شده است.

(الف) نمونه هایی از DDI های مثبت با چندین موضوع (داروها) به یک شی. در این سناریو، سوژه ها هیچ گونه برهمکنشی با یکدیگر ندارند، بلکه فقط با شیء برهمکنش دارند.

(ب) یک دارو را به عنوان موضوع و دو دارو را به عنوان شیء نشان می دهد. در این مورد، مطمئناً ارتباطی بین سوژه و ایزه به صورت جداگانه و با خوشه ای از اشیاء وجود دارد، اما هیچ برهمکنشی بین داروهای شیء وجود ندارد.

۲. بررسی پژوهش

در این بخش، کار تحقیقاتی اخیر را در رابطه با دیدگاه‌های مختلف وظیفه استخراج DDI، از جمله چالش‌ها، جنبه‌های مختلف و استراتژی‌های استخراج DDI، که توسط محققان در سال‌های اخیر مورد بررسی قرار گرفته‌اند، مورد بحث قرار خواهیم داد. ثانیاً، ما در مورد یادگیری عمیق در حوزه استخراج و طبقه‌بندی DDI در بخش ۳-۲ بحث می‌کنیم و به کارگیری تعداد کمی از آثار تحلیل احساسات در حوزه زیست پزشکی را بررسی می‌کنیم.

۲-۱. کار تداخل دارویی

با توجه به پیشرفت‌های زیست‌پزشکی، اصطلاحات جدید متعددی در حال تکامل هستند و ارتباط بین آنها از اهمیت بالایی برخوردار است. از آنجایی که این روابط معمولاً در متن زیست پزشکی ذکر شده است که شامل حجم بسیار زیادی است و با سرعت بسیار بالایی در حال رشد است [۳۴]، این حوزه جذابیت زیادی از سوی جامعه علمی دریافت می‌کند. وظیفه استخراج DDI به اندازه برهمکنش ژن-ژن و برهمکنش پروتئین - پروتئین مورد توجه قرار نگرفت. کار بسیار کمتری در استخراج خودکار DDIها وجود دارد، و در ابتدا این کار توسط [۳۵]، [۳۶] انجام شد، اما این کار به درستی استخراج نشد، به این معنی که از رویکردهای منسوخ در ارتباط با رویکردهای مدرن امروزی در جهان استفاده کرد. محققان نه تنها تکنیک‌هایی را در استخراج خودکار DDIها پیشنهاد کردند، بلکه یک مجموعه مشروح استاندارد شده برای ایجاد انگیزه در پژوهشگران در این کار خاص ارائه کردند. این مجموعه DDI توسط [۳۷] در سال ۲۰۱۳ ایجاد شد و شامل ۱۸۵۰۲ ماده دارویی مشروح دستی و همچنین ۵۰۲۸ جفت DDI است. در دسترس بودن مجموعه با چالش‌هایی همراه است. روش‌های مختلفی برای استخراج رابطه از ادبیات پزشکی پیشنهاد شده‌اند و تکنیک‌های مبتنی بر یادگیری ماشین (ML) همچنان در ساخت چنین سیستم‌هایی غالب هستند [۳۸]. این سیستم‌ها عمدتاً مبتنی بر طبقه‌بندی هستند و هر جفت دارو-دارو به یک مورد احتمالی از تداخل یا موارد دیگر طبقه‌بندی می‌شود. در انجام این کار، پیش پردازش زبان به شدت برای تبدیل داده‌های متنی به یک نمایش ساختاریافته استفاده می‌شود.

تحت SemEval-2013 (وظیفه ارزیابی معنایی)، یک چالش باز برای استخراج و طبقه‌بندی نام داروها اعلام شد. بسیاری از محققان این چالش را پذیرفتند و روش‌ها و رویکردهای مختلفی را اجرا کردند، اما بهترین رویکرد یادگیری مبتنی بر CRF بود [۳۹]، که در آن از «مواد شیمیایی مورد علاقه زیست‌پزشکی» معروف به ChEBI استفاده شد و به یک اندازه‌گیری F از آن دست یافتند. ۵۷٪ برای مجموعه داده کلی. شایان ذکر است که این چالش امتیاز ۰/۷۱ را در برابر F-Measure هدف قرار داده است. به طور مشابه، مسابقه دیگری برای استخراج DDI در سال ۲۰۱۳ برای رسیدگی به مسائل استخراج DDI برگزار شد. در مقابل چالش DDIExtraction-2013، چهار مورد از برجسته‌ترین موارد ارسالی وجود دارد. در این میان، روش مبتنی بر ماشین بردار پشتیبان (SVM-) از سایر رویکردها بهتر عمل کرد [۴۱]، [۴۳]. این روش‌ها عمدتاً شامل ویژگی‌های دستی هستند که در سطح کلمه کار می‌کنند و رسیدگی به جملات طولانی با نمایش روابط پیچیده همانطور که مجموعه MEDLINE-2013 شامل آن است، بسیار دشوار است.

۲-۲. یادگیری عمیق

از چند سال گذشته، برخلاف روش‌های مبتنی بر SVM، با تکامل یادگیری عمیق [۴۴]، روش‌های مبتنی بر شبکه‌های عصبی مانند شبکه‌های عصبی کانولوشنال (CNN) یا شبکه‌های عصبی تکراری (RNN) برای کارهای طبقه‌بندی مختلف استفاده می‌شوند. و کاملاً قوی و همه‌کاره هستند. CNN و RNN به ویژگی‌های دستی برای استخراج مقید نیستند و بنابراین در بسیاری از وظایف NLP مانند تجزیه و تحلیل احساسات [۴۵]، بازیابی پرس و جوی جستجو [۴۶] و تجزیه معنایی سریع و دقیق [۴۷] عملکرد بالاتری دارند. روش‌های مبتنی بر CNN نیز برای کار استخراج DDI آزمایش شده‌اند [۴۸]، [۵۱]. این روش‌ها نسبت به مدل‌های مبتنی بر SVM عملکرد خوبی داشتند. با این حال، عملکرد سیستم‌ها با مجموعه داده ML-

۲۰۱۳ رضایت بخش نبود. دلیل نتایج نامطلوب CNNها در ML-2013 این است که شامل ساختارهای طولانی و پیچیده جملاتی است که موقعیت رابطه حاوی کلمات ثابت نیست (یعنی قبل، بعد یا وسط داروها). بنابراین، برخی از کلمات معنی دار در جملات طولانی نادیده گرفته می شوند که در نهایت CNNها را در این زمینه ضعیف می کند.

برای غلبه بر این مشکل، سیستم های مبتنی بر RNN، برای طبقه بندی روابط استفاده می شود. در این سیستم ها، مکانیسم مدل حافظه کوتاه مدت بلند مدت (LSTM) ترکیب شده است که بر روی واحدهای حافظه (MUS) کار می کند. در LSTM، تمام اطلاعات متنی کلمات، خواه متوالی یا غیر متوالی، بدون در نظر گرفتن جملات دوردست به راحتی در MUها تقلید می شوند. این روش از دنباله ای برای پردازش کلمات یک جمله در یک RNN پیروی می کند، احتمال کمی از دست دادن اطلاعات برای وابستگی روابط در الگوهای راه دور وجود دارد. این شماره RNN همچنین به دلیل تسلط بر کلمات بعدی در پردازش الگوهای طولانی مانند ML-2013، آن را به یک شبکه مغرضانه تبدیل می کند. برای کاهش این موضوع، مکانیزم (توجه) مورد سوء استفاده قرار می گیرد که در آن روش وزن دهی برای اختصاص امتیاز به بخش های متن گنجانده شده است. به این ترتیب، بخش های با امتیاز بالا به عنوان بخش تأثیرگذارتر جمله در نظر گرفته می شوند و نسبتاً بیشتر مورد توجه شبکه قرار می گیرند. یک مکانیسم لایه مبتنی بر توجه به طور کاملاً مؤثر اعمال می شود و در [۵۴]، [۵۶] هنگامی که در ترجمه ماشینی مختلف و وظایف پاسخگویی به سؤال به کار می رود، نتایج بهتری به دست می آورد. این روش های مبتنی بر [۵۷]، [۵۹] RNN در وظیفه استخراج DDI به کار می روند و نتایج ۲ تا ۳ درصد بهتر از روش های پایه موجود با استفاده از SVM و شبکه های عصبی هستند، اما هنوز در فراخوان کم هستند. برای غلبه بر مسئله تعصب RNN، ما یک مدل RNN مبتنی بر LSTM با دو لایه اضافی مجزا، یعنی RNN پایین (RNNB) و RNN بالا (RNNT) پیشنهاد می کنیم. کار دقیق این روش (SEV_DDI) در بخش ۳-۳ مورد بحث قرار گرفته است.

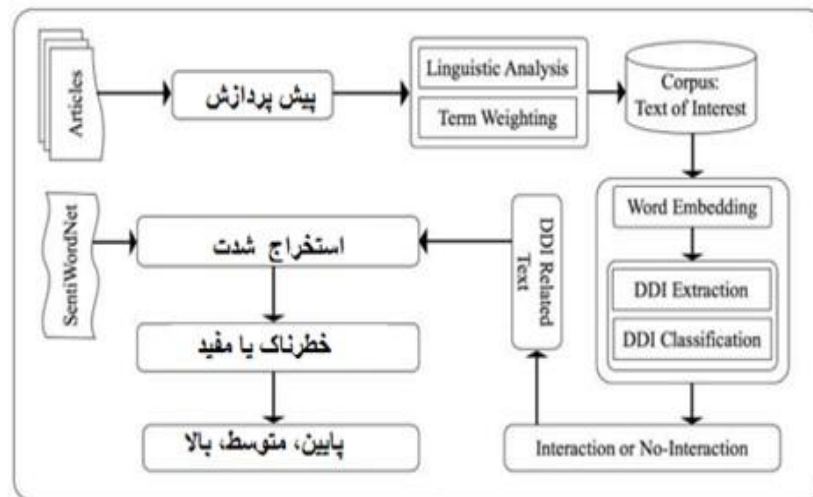
۳-۲. پیش بینی شدت در تداخلات دارو-دارو

وظیفه تجزیه و تحلیل احساسات یک حوزه تحقیقاتی ثابت است که معمولاً بر روی متن عقیده جمع آوری شده از رسانه های اجتماعی اعمال می شود. (احساس) می تواند منعکس کننده جدیت، شدت یا قطعیت یک برهمکنش گزارش شده در متن پزشکی باشد. در حوزه زیست پزشکی، تا جایی که ما می دانیم، تجزیه و تحلیل احساسات در جنبه های بسیار کمی اعمال می شود. نویسندگان [۶۰] با کار بر روی ادبیات مراقبت های بهداشتی با توجه به تحلیل احساسات، قطبیت را در سطح جمله با استفاده از ویژگی های زبانی چندگانه پردازش متن، به عنوان مثال، n-gram بررسی و طبقه بندی کردند. در این کار، نیو و همکاران. [۶۱] از یک سیستم زبان پزشکی یکپارچه (UMLS) برای استخراج اطلاعات زمینه و دسته بندی اصطلاحات پزشکی استفاده می کنند. به منظور استخراج احساسات از متن پزشکی، یک مکانیسم چند مرحله ای طراحی شد که در آن ابتدا موضوعات یا موضوعات طبقه بندی شدند و سپس قطبیت هر موضوع بررسی شد [۶۲] که ۱۰ درصد عملکرد بهتری را نشان می دهد.

نویسندگان [۶۳] که در حوزه تداخلات احتمالی دارو-دارو کار می کنند، ۹۲ مقاله منتشر شده و تقریباً ۱۲ مصاحبه را برای استخراج ۵۶ جنبه مختلف از دانش برای تفسیر DDIها بررسی کرده اند. در این بررسی، نویسنده پیشرفت ها و نقاط قوت سیستم های خبره را به همراه محدودیت ها و نقاط ضعف این سیستم های خبره شناسایی کرده است. استخراج شدت DDIها به ندرت مورد مطالعه قرار می گیرد، بنابراین ما همچنین روش استخراج شدت را با استفاده از قطبیت کلمات توصیف کننده برهمکنش پیشنهاد می کنیم. وظیفه استخراج شدت DDI مورد توجه زیادی قرار نگرفت، اما می تواند برای پیشرفت های سیستم های پشتیبانی تصمیم بالینی بسیار امیدوارکننده باشد. به جای استخراج صرف رویداد DDI، شدت یا اهمیت تداخل دارو بسیار حیاتی است. در نتیجه، ما مکانیزمی طراحی کردیم که شدت DDI طبقه بندی شده را نیز استخراج می کند.

۳. روش ها و اجرا

ما یک مدل مبتنی بر شبکه عصبی عمیق با مکانیزم سلسله مراتبی برای استخراج DDI و شدت جملات متنی که از ادبیات زیست پزشکی (یعنی انتشارات) جمع‌آوری شده‌اند و با استفاده از تکنیک‌های NLP روی آنها پردازش می‌شوند، پیشنهاد می‌کنیم. این مدل از RNN با افزودن لایه‌های سلسله‌مراتبی LSTM استفاده می‌کند. یک روش مبتنی بر واژگان نیز برای بررسی شدت برهمکنش بین DDI طراحی شده است. در این بخش، مراحل مربوط به مدل مبتنی بر RNN را با مکانیزم سلسله‌مراتبی نشان می‌دهیم.



شکل ۲- معماری سیستم پیشنهادی SEV-DDI

۳-۱. پیش پردازش

عملیات پیش پردازش متداول و پاکسازی متن روی جملاتی انجام می‌شود که شامل واژه سازی است، اما محدود به آن نیست. هر دارویی که در یک جمله ذکر شده است برچسب گذاری می‌شود و با داروهای دیگر تداخل دارد. تعداد جفت داروها (DP) در یک جمله را می‌توان با رابطه (۱) تخمین زد.

$$Drug\ Pairs\ (DP) = \max\left(0, \sum_{i=1}^n (i-1)\right) \quad (1)$$

n تعداد دارو در یک جمله است

کور کردن دارو نیز اعمال می‌شود، جایی که نام هر دارو با برچسبی مطابق با [۶۴] اختصاص داده می‌شود، به عنوان مثال، برای جمله «آسپرین ممکن است اثرات پروبنسید را کاهش دهد»، جمله برچسب‌گذاری شده «دارو^A ممکن است اثرات دارو^B را کاهش دهد» است. تکنیک کور کردن دارو به یک مدل کمک می‌کند تا این برچسب‌ها را به عنوان «موضوع» و «هدف» تشخیص دهد که در نهایت به مدل در طبقه‌بندی کمک می‌کند. سپس این جملات پردازش شده برای شناسایی و طبقه‌بندی DDIها به مدل ارائه می‌شود.

۳-۲. معماری شبکه عصبی تکراری

مدل شبکه عصبی بازگشتی پیشنهادی ما، با حافظه کوتاه مدت سلسله مراتبی (LSTM)، شامل پنج لایه است. این لایه ها شامل «جاسازی کلمه»، «توجه»، «ترک کردن»، «RNN پایین» و «RNN بالا» همانطور که در شکل ۳ نشان داده شده است. عملکرد و هدف هر لایه در مدل مورد نظر در ادامه مورد بحث قرار گرفته است.

لایه جاسازی کلمه

در این لایه، هر کلمه به یک بردار با ارزش واقعی تبدیل می شود. این نگاشت کلمات در ماتریس با استفاده از Word2Vec از قبل موجود و جاسازی اطلاعات [۶۶] با استفاده از چکیده های PubMed حاوی داروها [۶۷] انجام می شود. هر جمله از پیش پردازش شده است و (si) و (dj) را تشکیل می دهد که در آن dj یک برچسب دارو است و si هر کلمه دیگری در این جمله خاص است. هر کلمه (si) با استفاده از ماتریس تعبیه کلمه با فرمول زیر به بردار کلمه تبدیل می شود. WEMB یک ماتریس جاسازی است و $|V| \times |WEMB Rds|$ ، که در آن d_s تعداد ابعاد در جاسازی کلمه و V واژگان در مجموعه داده های آموزشی است، v_i^s در مقابل معادله (۲) نمایانگر شاخص جاسازی کلمه است.

$$s_i^{\rightarrow} = W_{EMB} \cdot v_i^s \quad (2)$$

لایه توجه

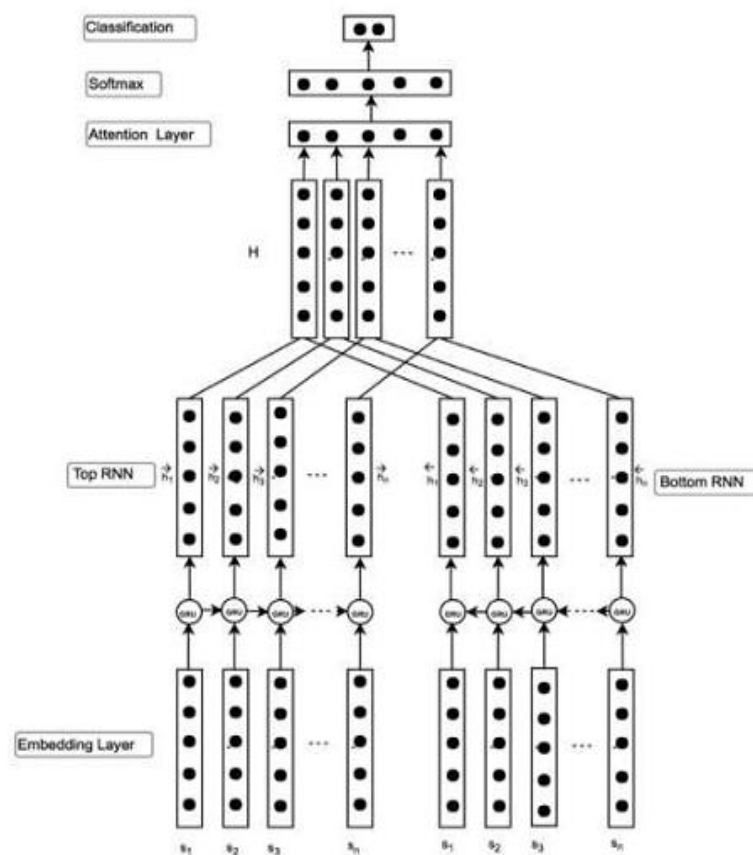
کلمات بین و اطراف فاعل و ابژه برای تشخیص اینکه آیا کنش متقابل یا تأثیر موضوع بر شی تأثیر می گذارد بسیار مهم هستند. بنابراین، نه همه، بلکه تعداد کمی از این کلمات مهم (مانند افزایش، کاهش، تأثیر، نظارت دقیق، توصیه، ممنوعیت، و غیره) در تشخیص تداخل بین داروها ارزش زیادی دارند. بنابراین، ما از یک لایه توجه پس از جاسازی کلمات در بردارها استفاده می کنیم زیرا لایه توجه به طور غالب بر روی مکانیسم یادگیری جمله به جمله کار می کند. به طور مشابه، برای وظیفه استخراج رابطه [۶۸] اعمال شده است. بنابراین، این مکانیسم در مدل ما نیز همانطور که در [۶۹] اعمال شد با هدف ارائه اطلاعات برداری تعبیه شده تصفیه شده تر به شبکه اعمال می شود.

لایه حذفی

الگوریتم های مبتنی بر یادگیری ماشین در حین پردازش حجم عظیمی از داده ها با ابعاد بسیار زیاد، بیش از حد برازنده می شوند. تکنیک (Dropout) توسط [۷۰] آغاز شد، که با حذف مقادیر ورودی که می تواند باعث این مشکل شود، از برازش بیش از حد مدل جلوگیری می کند. بنابراین، لایه داخلی بیشتر ورودی های متمایز را پردازش می کند. برای جلوگیری از برازش بیش از حد این مدل، ما این استراتژی را در واحد بازگشتی دروازه ای (GRU) با احتمال در آموزش تعبیه سازی و آموزش شبکه عصبی با مقادیر تست چندانگانه اعمال کردیم.

لایه های RNN پایین و بالا

دو لایه مبتنی بر RNN برای غلبه بر مشکل تعصب RNN ها با LSTM استفاده می شود. لایه های بالا و پایین به طور جداگانه اعمال می شوند (مانند کار دوطرفه LSTM) برای ایجاد یک امتیاز از هر دو انتها و سپس ادغام می شوند تا نمره نهایی کلمات با مشکل سوگیری صفر تولید شود. RNN ها می توانند هر نوع الگوی جمله ای را از طریق GRU خود پردازش کنند که در آن هر کلمه (s_i) از جمله (S) به همراه یک حالت پنهان قبلی (h_{i-1}) به عنوان ورودی گرفته می شود. شبکه LSTM [۷۲] برای کاهش موضوع رشد نمایی و فروپاشی بردارهای گرادیان در الگوهای طولانی معرفی شد که نتایج بهتری را نشان داد [۷۲]. LSTM همچنین زمانی که همراه با RNN به کار می رود، به عنوان یک شبکه مغرضانه شناخته می شود، زیرا در این تسلط شبکه، کلمات اولیه یک جمله نادیده گرفته می شوند و کلمات بعدی به تدریج از کلمات اولیه غالب تر می شوند [۶۹].



شکل ۳- معماری شبکه عصبی عمیق پیشنهادی

برای غلبه بر این تعصب، ما یک مکانیسم دو لایه طراحی کردیم، که در آن LSTM از هر دو انتهای یک جمله آغاز می شود تا اطلاعات متنی آن را همراه با اطلاعات معنایی جهانی ضبط کند. هنگام پردازش یک جمله $S_i = \{s_1, s_2, s_3, s_4, \dots, s_n\}$ ما دو ماتریس کدگذاری شده همانطور که در معادلات (۳) و (۴) نشان داده شده است. در پایان، بردارهای خروجی هر دو LSTM در معادله (۵) به هم پیوسته اند تا اطلاعات متنی و معنایی یک جمله را به دست آورند.

$$H_{S_i}^{\rightarrow} = \{h_1^{\rightarrow}, h_2^{\rightarrow}, h_3^{\rightarrow}, \dots, h_{n-1}^{\rightarrow}, h_n^{\rightarrow}\} \quad (1)$$

$$H_{S1}^{\leftarrow} = \{h_n^{\leftarrow}, h_{n-1}^{\leftarrow}, h_{n-2}^{\leftarrow}, \dots, h_2^{\leftarrow}, h_1^{\leftarrow}\} \quad (2)$$

$$H = \{H_{S1}^{\rightarrow} \parallel H_{S1}^{\leftarrow}\} \approx \{h_1^{\rightarrow} \parallel h_1^{\leftarrow}, h_2^{\rightarrow} \parallel h_2^{\leftarrow}, h_3^{\rightarrow} \parallel h_3^{\leftarrow}, \dots, h_n^{\rightarrow} \parallel h_n^{\leftarrow}\} \quad (3)$$

هر جفت داروی «واقعی» در میان یکی از چهار نوع، یعنی مکانیسم، اثر، هدف و توصیه طبقه‌بندی می‌شود. شرح نوع دارو به همراه نمونه‌های مربوط به آنها در جدول ۱ نشان داده شده است. تمام مثال‌های جمله از مجموعه داده DDIExtraction-2013 گرفته شده‌اند.

جدول ۱- شرح انواع DDI در DDIExtraction-2013

نوع DDI	شرح	جمله
Int	این برهمکنش در یک جمله گزارش شده است اما اطلاعات دقیقی در مورد برهمکنش ارائه نشده است.	مصرف همزمان الکل با فنتریمین هیدروکلراید ممکن است منجر به تداخل دارویی شود.
Advise	برای یک جفت دارو، توصیه‌هایی در مورد استفاده از آن وجود دارد.	اسکوپولامین باید با احتیاط در بیمارانی که از داروهای دیگری استفاده می‌کنند که می‌توانند اثرات CNS مانند آرام بخش‌ها، آرام بخش‌ها یا الکل را ایجاد کنند، استفاده شود.
Mechanism	بین یک جفت دارو، مکانیسم فارماکوکینتیک وجود دارد.	سطح پنی سیلین خون ممکن است با تجویز همزمان پروبنسید که ترشح لوله‌های کلیوی پنی سیلین‌ها را مسدود می‌کند طولانی شود.
Effect	بین یک جفت دارو، تأثیری در جمله گزارش می‌شود که می‌تواند مثبت یا منفی باشد.	سایر مهارکننده‌های HDAC ترومبوسیتوپنی شدید و خونریزی گوارشی با مصرف همزمان ZOLINZA و سایر مهارکننده‌های HDAC (مانند والپروئیک اسید) گزارش شده است.
False	هیچ تداخلی بین دو داروی ذکر شده وجود ندارد	اتصال درون تنی وارفارین به پروتئین‌های پلاسمای انسانی تحت تأثیر تولمتین قرار نمی‌گیرد و تولمتین زمان پروترومبین را در داوطلبان عادی تغییر نمی‌دهد.

۳-۳. روش استخراج شدت

ما فقط به جملاتی علاقه‌مندیم که حداقل یک نمونه DDI برای آنها (درست) باشد. پس از طبقه‌بندی جفت‌های دارویی برهمکنشی، جملات گزارش DDI از مجموعه مشروح استخراج می‌شوند [۲۲]. همانطور که در شکل ۲ ذکر شد، جملات استخراج شده برای محاسبه احساسات، یعنی قطبیت کلمات، از پیش پردازش می‌شوند. در پیش پردازش، جایی که نام داروها (به عنوان مثال، اسامی) از متن مربوطه حذف می‌شود زیرا نام داروها صرفاً حاوی اسم‌های قطبی است و نمی‌تواند در رتبه‌بندی قطبیت موثر باشد. ما متن را فقط با حفظ افعال اصلاح کردیم. قیدها و صفت‌ها و اسامی حذف می‌شوند تا سیستم ما از انجام پردازش اضافی جلوگیری کند.

واژگانی مانند Sent WordNet و WordNet Affect واژگان عمومی هستند که برای استخراج احساسات عمومی متن، به عنوان مثال، فیلم‌ها و نقدهای اجتماعی استفاده می‌شوند. واژگان ذهنی [۷۰] برای استخراج عبارت ذهنی از استدلال‌ها یا

گزاره های متنی استفاده می شود. بسیاری از واژگان عمومی و ذهنی در تحقیقات پزشکی برای وظایف مختلف مراقبت های بهداشتی اقتباس شده اند. واژگان دارویی توسعه یافته نیز به طور خاص برای حوزه های مراقبت های بهداشتی و زیست پزشکی تکامل یافته است که برای استخراج احساسات متن دارویی و بالینی استفاده می شود. ما قطبیت جملات را با استفاده از Sent WordNet استخراج کردیم و برهمکنش در سطوح پایین، متوسط یا بالا برای DDI های خطرناک و همچنین سودمند براساس قطبیت جمله نامزد طبقه بندی می شود. تجزیه و تحلیل پیامدهای شدت در بخش ۱-۴ نشان داده شده است.

۴. داده ها، پارامترهای تجربی و نتایج

مجموعه [DDIExtraction-2013 22] که با نسخه های بانک دارو و چکیده مقالات PubMed در مجموعه MEDLINE ذخیره می شود. این شامل ۱۸۵۰۲ ماده دارویی همراه با ۵۰۲۸ جفت DDI است. آمار مجموعه داده در جدول ۲ نشان داده شده است.

جدول ۲- آمار DDIExtraction-2013

	DrugBank			MEDLINE		
مطالب	آموزشی	آزمایشی	کل	آموزشی	آزمایشی	کل
مقالات	۵۷۲	۱۵۸	۷۳۰	۱۴۲	۳۳	۱۷۵
جفت دارو	۲۶,۰۰۵	۵۲۶۵	۳۱,۲۷۰	۱۷۸۷	۴۵۱	۲۲۳۸
Positive DDI Pairs	۳۷۸۹	۸۸۴	۴۶۷۳	۲۳۲	۹۵	۳۲۷
Negative DDI Pairs	۲۲,۲۱۶	۴۳۸۱	۲۶,۵۹۷	۱۵۵۵	۳۵۶	۱۹۱۱

تعداد صفحه مقاله حداکثر ۱۲ صفحه است. مقاله باید به نمایش مشروح این مجموعه با جملات و تمام موارد ممکن، یعنی جفت دارو، در شکل ۴ نشان داده شده است. برای یک جمله، همه موارد ممکن است (نادرست) باشند، زیرا یک جفت دارو در شکل ۴ با یک بلوک نقطه دار سیاه نشان داده شده است. در مقابل، یک برهمکنش گزارش شده بین یک جفت دارو با (درست) نشان داده شده است، همانطور که در شکل ۴ با یک بلوک مستطیل قرمز نشان داده شده است. ممکن است بیش از یک مثال «درست» و «نادرست» برای یک جمله وجود داشته باشد و همه می توانند درست یا نادرست باشند.

```
<sentence id="DDI-DrugBank.d747.s0" text="Tetracycline, a bacteriostatic
antibiotic, may antagonize the bactericidal effect of penicillin and
concurrent use of these drugs should be avoided.">
  <entity id="DDI-DrugBank.d747.s0.e0" charOffset="0-11"
    type="group" text="Tetracycline"/>
  <entity id="DDI-DrugBank.d747.s0.e1" charOffset="31-40"
    type="group" text="antibiotic"/>
  <entity id="DDI-DrugBank.d747.s0.e2" charOffset="85-94"
    type="drug" text="penicillin"/>
  <pair id="DDI-DrugBank.d747.s0.p0" e1="DDI-DrugBank.d747.s0.e0"
    e2="DDI-DrugBank.d747.s0.e1" ddi="false"/>
  <pair id="DDI-DrugBank.d747.s0.p1" e1="DDI-DrugBank.d747.s0.e0"
    e2="DDI-DrugBank.d747.s0.e2" ddi="true" type="effect"/>
  <pair id="DDI-DrugBank.d747.s0.p2" e1="DDI-DrugBank.d747.s0.e1"
    e2="DDI-DrugBank.d747.s0.e2" ddi="false"/>
</sentence>
```

شکل ۴- حاشیه نویسی جملات در مجموعه DDI

ما از این مجموعه برای آموزش و تست مدل استفاده کردیم. اسناد در ابتدا از پیش پردازش شده و در فایل های ترشی برای آموزش و آزمایش مدل ذخیره می شدند. پارامترهای آزمایش طبقه بندی DDI در جدول ۳ نشان داده شده است. بقیه پارامترهایی که در جدول نشان داده نشده اند به عنوان تنظیمات پیش فرض مانند TensorFlow نگهداری می شوند.

جدول ۳- پارامترهای تنظیم

پارامتر	مقدار
ابعاد جاسازی کلمه	۱۰۰
ابعاد جاسازی موقعیت	۲۰
اندازه دسته کوچک	۶۰
ابعاد حالت پنهان	۲۳۰
کوتاه ترین طول مسیر	۱۲
میزان یادگیری	۰/۰۰۵
جاسازی حذف	۰/۷
حذف متراکم	۰/۵

دقت، یادآوری و اندازه گیری f برای کار استخراج DDI استفاده می شود و به ترتیب با فرمول های (۶) - (۸) محاسبه می شود. نتایج طبقه بندی DDI و شناسایی شدت در جدول ۴ نشان داده شده است و مقایسه با سیستم های پیشرفته نیز ارائه شده است.

$$Precision (P) = \frac{TP}{TP + FP} \quad (۶)$$

$$Recall (R) = \frac{TP}{TP + FN} \quad (۷)$$

$$F1 Measure (F1) = \frac{2 (Precision \times Recall)}{Precision + Recall} \quad (۸)$$

جدول ۴- مقایسه عملکرد با سیستم های موجود در DDIExtraction-2013

روش	سیستم	سال	دقت	فراخوانی (R)	F-Score (F1)
SVM-based methods	UTurku [30]	2013	73.2	49.9	59.4
	NIL_UCM [63]	2013	55.0	53.0	54.0
	SCAI [29]	2013	55.0	39.0	46.0
	UWM-TRIADS [64]	2013	43.0	50.0	47.0
	UColorado_SOM [65]	2013	27.0	43.0	33.0
	BioSem [66]	2014	67.0	52.0	59.0
	FBLKA [28]	2015	nil	nil	67.0
	Raihani and Laachfoubi [67]	2016	73.7	68.7	71.1
	Zheng et al. [34]	2016	nil	nil	68.4
Neural Nets-based methods	MCCNN-DDI [40]	2016	nil	nil	67.8
	SCNN [41]	2016	68.5	61.0	64.5
	CNN-DDI [37]	2016	75.3	60.4	67.0
	Joint-LSTMs [39]	2017	71.3	66.9	69.3
	RHCNN [69]	2019	77.3	73.75	75.48
	SGRU-CNN [71]	2020	76.19	73.34	74.74
	AGCN [72]	2020	78.17	75.59	76.86
	Our Method: SEV-DDI	2021	<u>83.81</u>	<u>81.59</u>	<u>82.68</u>

در نتیجه، وظیفه استخراج شدت از متن DDI اعمال می شود. مکانیسم پیش بینی شدت با پیاده سازی الگوریتم ۱ و اعمال آن بر روی مجموعه داده های DDIExtraction-2013 اعمال می شود. ما ۴۶۷۲ جمله از مجموعه DrugBank و ۳۲۷ جمله از MEDLINE شامل ۹۰۰۰ داروی مجزا و ۴۹۹۹ DDI جمع آوری کردیم. در الگوریتم، Polpos و Polneg به ترتیب نشان دهنده قطبیت مثبت و منفی کلمات هستند. پس از جمع تمام قطبیت های کلمات، قطبیت نهایی جمله با کم کردن Polneg از Polpos محاسبه می شود. بر اساس قطبیت نهایی جمله، هر نمونه DDI با یک برجسب شدت تخصیص داده می شود، همانطور که در الگوریتم ۱ توضیح داده شده است. عملکرد سیستم در جدول ۴، تصویر شدت در جدول ۵ نشان داده شده است، در حالی که جدول ۶ آمار طبقه بندی شده DrugBank و پیکره MEDLINE با درجه شدت استخراج شده در متن ارائه شده است.

الگوریتم: پیش بینی شدت در برهمکنش دارو-دارو

- ۱: متن برای هر نمونه DDI تقسیم می شود. به هر جمله جداگانه پرداخته می شود.
- ۲: هر جمله با استفاده از کیت ابزار NLTK از قبل پردازش و نرمال سازی شده است.
- ۳: هر جمله با استفاده از SentiWordNet پردازش می شود و قطبیت کلمات استخراج شده و به ترتیب برای قطب مثبت و منفی به Pol^{pos} و Pol^{neg} خلاصه می شود.
- ۴: قطب نهایی به صورت زیر محاسبه می شود: قطب نهایی = $Pol^{pos} - Pol^{neg}$
- ۵: قطب نهایی به دو مجموعه تقسیم می شود، یعنی سودمند و خطرناک:
- ۶: اگر قطب نهایی ≤ 0 باشد، پس:
- ۷: نمونه DDI سودمند است.
- ۸: دیگری
- ۹: نمونه DDI خطرناک است.
- ۱۰: پایان اگر
- ۱۱: طبقه بندی موارد مفید و خطرناک به سطوح شدت.
- ۱۲: برای هر نمونه در سودمند:
- ۱۳: اگر قطب نهایی ۰ تا ۰/۲ باشد، DDI کمی سودمند است.
- ۱۴: در غیر این صورت اگر قطب نهایی ۰/۳ تا ۰/۶ باشد، DDI نسبتاً سودمند است.
- ۱۵: در غیر این صورت DDI بسیار سودمند است.
- ۱۶: پایان اگر
- ۱۷: پایان برای
- ۱۸: برای هر نمونه در Dangerous:
- ۱۹: اگر قطب نهایی ۰ تا ۰/۲- باشد، DDI کمی خطرناک است.
- ۲۰: در غیر این صورت اگر قطب نهایی ۰/۳- تا ۰/۶- باشد، DDI نسبتاً خطرناک است.
- ۲۱: در غیر این صورت DDI بسیار خطرناک است.
- ۲۲: پایان اگر
- ۲۳: پایان برای

۴-۱. نتایج و بحث

عملکرد مدل با مدل های پیشرفته برای کار استخراج DDI در جدول ۴ مقایسه شده است. روش ما در هر سه معیار دقت، یادآوری و امتیاز f عملکرد بهتری داشت. هر یک از این سه معیار در کارهای مربوط به بیوپزشکی و به ویژه در استخراج DDI بسیار مهم هستند، مدل ما در هر سه معیار، به ترتیب ۸۳/۸۱، ۸۱/۵۹، ۸۲/۶۸ در دقت، یادآوری و اندازه گیری f بیش از ۸۰ درصد امتیاز کسب کرد. تا جایی که ما می دانیم، هیچ یک از سیستم های دیگر در آخرین نقطه زمانی به این نقطه عطف دست پیدا نکردند.

بهبود عملکرد کلی در کار DDI-Extraction، همانطور که در جدول ۴ ذکر شد، به دلیل اضافه شدن مکانیسم توجه و LSTM دو طرفه است. مکانیسم توجه ویژگی های بهتری را در آموزش ارائه می کند که در بیشتر فعل و انفعالات دخیل بود، و تسلط این کلمات با توجه به هر داروی کاندید نیز افزایش می یابد. در نتیجه، استفاده از LSTM از هر دو انتهای جمله، مسئله سوگیری را حل کرد که در کار استخراج DDI که در آن مهم ترین کلمات می توانستند در هر جایی از جمله باشند، یعنی قبل،

بعد یا اطراف موضوع، حیاتی بود. هر کلمه در یک جمله بدون در نظر گرفتن قرار دادن آن در جمله به تسلط یکسانی دست یافت که به پیش بینی دقیق تری در استخراج و طبقه بندی DDI دست یافت، همانطور که نتایج نشان داده شد. در کار استخراج شدت، تمام جملات مربوط به DDI در سطوح خفیف، متوسط یا شدید برهمکنش دسته بندی می شوند. جدول ۵ دسته بندی DDI ها را در سطح یا درجه ای از شدت نشان می دهد که در آن نمونه هایی از DDI Extraction- 2013 با درجه شدت آنها در برهمکنشی بر اساس قطبیت جملات نشان داده شده است. احساسات مثبت در جمله مربوط به DDI دارای یک تداخل خفیف در نظر گرفته می شود که در آن داروشناسان یا کارشناسان توصیه می کنند برخی از داروها را با هم برای ایمنی یا درمان برخی بیماری ها مصرف کنند، در حالی که سطح متوسطی از شدت به آن دسته از DDI هایی اختصاص داده می شود که قطبیت آنها خنثی است، به عنوان مثال، که در آن کارشناسان برخی مکانیسم ها یا تأثیر ضعیف بر سایر داروها را شرح داده اند. سطح شدیدی از تداخل از احساسات منفی ناشی می شود که در آن کارشناسان بالینی خطر را نشان دادند و توصیه هایی در مورد نظارت دقیق ارائه کردند یا تأثیر قابل توجهی از یک دارو بر روی سایرین نشان دادند. آمار طبقه بندی DDI در جدول ۶ نشان داده شده است.

جدول ۵- طبقه بندی شدت DDI

سطح شدت	جفت دارو	جمله
بسیار سودمند	(نمک، پریندوپریل)	احتمال اثرات کاهش فشار خون را می توان با قطع مصرف دیورتیک یا افزایش مصرف نمک قبل از شروع درمان با پریندوپریل به حداقل رساند.
ذینفع متوسط	(AAV2، کندرویتین ABC لیاز)	انتقال شبکه با واسطه AAV2 با تزریق همزمان هپاریناز III یا کندرویتین ABC لیاز بهبود می یابد.
پایین	(اسکوپولامین، آرام بخش)، (اسکوپولامین، آرام بخش)، (اسکوپولامین، الکل)	اسکوپولامین باید با احتیاط در بیمارانی که از داروهای دیگری استفاده می کنند که می توانند اثرات CNS مانند آرام بخش ها، آرام بخش ها یا الکل را ایجاد کنند، استفاده شود.
پایین	(پانتوپرازول، کلپیدوگرل)	پانتوپرازول اثر بسیار ضعیف تری بر فارماکوکینتیک کلپیدوگرل و واکنش پلاکتی در طول مصرف همزمان دارد.
متوسط خطرناک	(الکل، فنترمین هیدروکلراید)	مصرف همزمان الکل با فنترمین هیدروکلراید ممکن است منجر به تداخل دارویی شود.
بسیار خطرناک	(مهارکننده های HDAC، ZOLINZA)	سایر مهارکننده های HDAC ترومبوسیتوپنی شدید و خونریزی گوارشی با مصرف همزمان ZOLINZA و سایر مهارکننده های HDAC (مانند والپروئیک اسید) گزارش شده است.

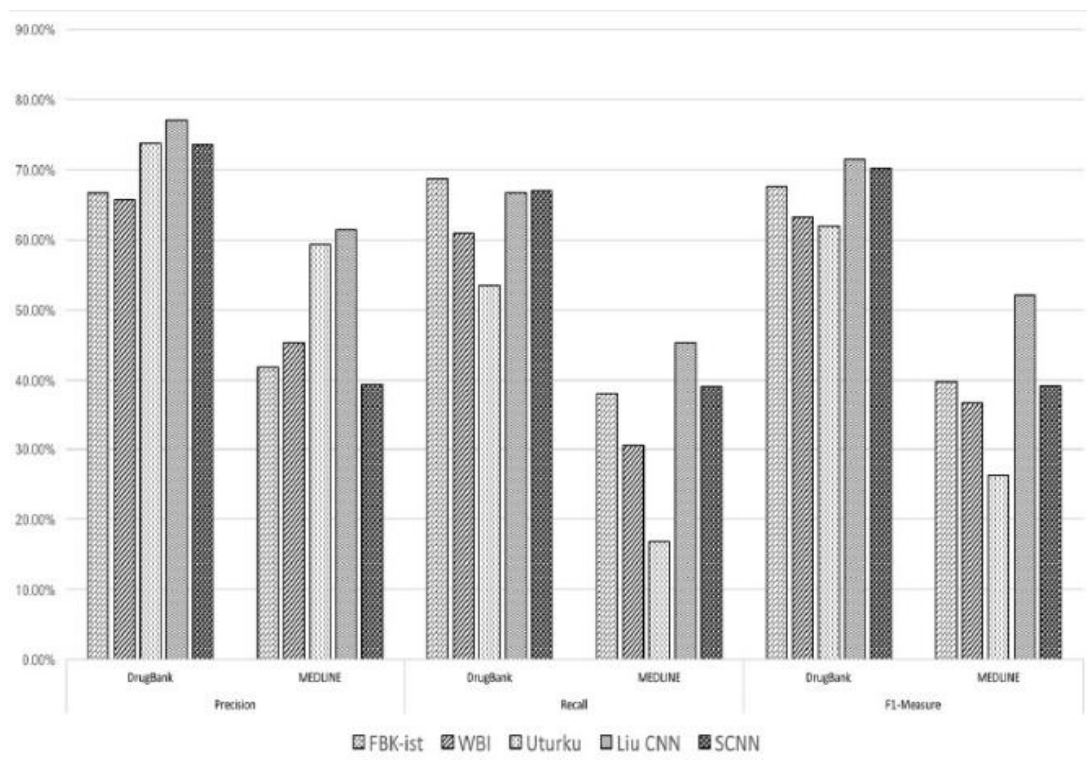
جدول ۶- پیش بینی شدت DDI در DrugBank و MEDLINE

کل	MEDLINE	DrugBank	
۹۰۵	۱۷۵	۷۳۰	Articles
۳۳/۵۰۸	۲۲۳۸	۳۱,۲۷۰	Candidate DDIs
۴۹۹۹	۳۲۷	۴۶۷۲	Positive DDI Related Sentence
۸۱۸	۸۹	۷۲۹	Low
۸۳۲	۷۳	۷۵۹	Moderate Beneficial DDIs
۸۰۱	۳۹	۷۶۲	High
۳۵۸	۲۲	۳۳۶	Low
۷۸۰	۵۶	۷۲۴	Moderate Dangerous DDIs
۱۴۱۰	۴۸	۱۳۶۲	High

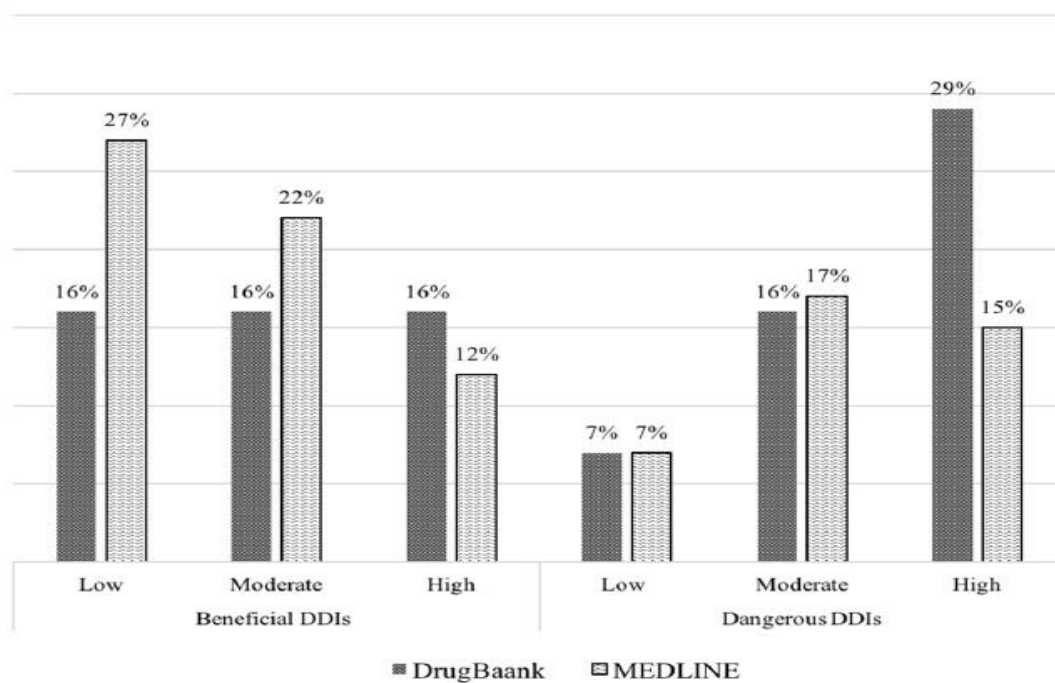
ما می‌توانیم ادعا کنیم که این طبقه‌بندی با تجزیه و تحلیل شکل ۵، که آمار عملکرد طبقه‌بندی DDI را توسط سیستم‌های موجود در بانک دارو و MEDLINE به طور جداگانه ارائه می‌کند، دقیق است. این به وضوح نشان می‌دهد که عملکرد همه مدل‌ها در MEDLINE در مقایسه با بانک دارو پایین است و هیچ یک از مدل‌های موجود عملکرد بهتری نسبت به DrugBank در MEDLINE نداشتند.

DrugBank مقالاتی با ابهام کمتر دارد و برهمکنش واضحی را در متن نشان می‌دهد. بسیاری از مدل‌هایی که قبلاً اجرا شده بودند، در مقایسه با زمانی که برای MEDLINE اعمال می‌شدند، عملکرد بسیار خوبی را در زمانی که فقط برای مجموعه DrugBank اعمال می‌شد، نشان دادند. در ارتباط این تاکید با شدت استخراج شده ما، سیستم ما ۳۶ درصد DDI‌های DrugBank را به عنوان بسیار برهمکنشی طبقه بندی کرد، در حالی که فقط ۲۷ درصد از DDI‌های MEDLINE دارای برهمکنش بالایی هستند.

برعکس، مقالات MEDLINE corpora کمتر رسا هستند و ساختار جملات بسیار طولانی و پیچیده‌ای در گزارش یک برهمکنش دارند. بسیاری از سیستم‌های پیشرفته عملکرد بسیار ضعیفی را در مجموعه MEDLINE در مقایسه با DrugBank نشان دادند. شکل ۶ تجزیه و تحلیل آماری مربوط به عملکرد مدل‌ها را بر روی مجموعه‌های بانک دارو و MEDLINE نشان می‌دهد. این به وضوح نشان می‌دهد که تمام مدل‌های پیاده سازی شده نتوانستند عملکرد بهتری در MEDLINE نسبت به DrugBank داشته باشند، حتی برخی از مدل‌ها عملکرد بسیار بدی را در MEDLINE نشان دادند، در حالی که نتایج عالی را در مجموعه دارو بانک نشان دادند.



شکل ۵- آمار سیستم های پیشرفته در DrugBank و MEDLINE



شکل ۶- پیش بینی شدت DDI در DrugBank و MEDLINE

این ثابت می کند که مقالات موجود در مجموعه MEDLINE دارای پیچیدگی زیادی در بیان رابطه DDI ها هستند. به همین دلیل است که همیشه برای هر مدلی عملکرد بهتر در مجموعه MEDLINE برای استخراج و طبقه بندی DDI ها کاملاً چالش برانگیز است.

به طور مشابه، مدل استخراج شدت ما نیز درصد بسیار کمتری از DDI با شدت بالا را در هر دو طرف در بدنه MEDLINE نسبت به DrugBank نشان داد، یعنی به ترتیب ۳۶ درصد و ۲۷ درصد در DrugBank و MEDLINE. از طرف دیگر، سطوح کم و متوسط شدت در پیش بینی با ۳۴ درصد و ۳۹ درصد از مجموعه MEDLINE، در مقایسه با پیش بینی های DrugBank که ۲۲ درصد و ۳۲ درصد از DDI را برای سطوح خفیف و متوسط شدت نشان می دهد، همچنان که در شکل ۶ نشان داده شده است، غالب باقی ماند.

پیش بینی درجه شدت روی مجموعه کلی [DDIExtraction-2013 22] در جدول ۷ نشان داده شده است. در رتبه بندی های استخراج شده ما، حدود ۴۴ درصد از DDI ها در DDIExtraction-2013 موارد بسیار برهمکنشی هستند، ۲۴ درصد از آنها برهمکنش کم و ۳۲ درصد از برهمکنش در سطح متوسط هستند. این نتایج ممکن است در منظر فعل و انفعالات دارویی کمی متفاوت باشد، زیرا این فعل و انفعالات براساس قطبیت جملاتی که برای گزارش یک برهمکنش یا پیامدهای برهمکنش مقالات استفاده شده است، دسته بندی می شوند.

جدول ۷- پیش بینی سطوح شدت در DDIExtraction-2013

سطح شدت	DrugBank	MEDLINE	DDIExtraction-2013
پایین	۲۲ درصد	۳۴ درصد	۲۴ درصد
متوسط	۳۲ درصد	۳۹ درصد	۳۲ درصد
زیاد	۳۶ درصد	۲۷ درصد	۴۴ درصد

پیش بینی و حاشیه نویسی شدت DDI توسط یک متخصص داروسازی و مراقبت های بهداشتی می تواند کمی با پیش بینی ما متفاوت باشد. با این حال، برای یک متخصص بسیار چالش برانگیز است که تعداد زیادی از جفت ها را در سطوح شدت دسته بندی کند. افزودن شدت DDI ها در سیستم های خبره مراقبت های بهداشتی بالینی و دارویی می تواند در تهیه نسخه های دقیق تر و دقیق تر بسیار امیدوار کننده باشد. همچنین می تواند در کاهش ADR ها، که یکی از دلایل اصلی هزینه های مراقبت های بهداشتی و شرایط اکتسابی بیمارستان و بستری مجدد است، بسیار سودمند باشد.

۵. نتیجه گیری

در این پژوهش، ما یک مدل شبکه عصبی LSTM مبتنی بر توجه سلسله مراتبی را برای بهبود عملکرد طبقه بندی وظایف برهمکنش دارو-دارو پیشنهاد می کنیم. مدل ما از همه روش های موجود در عملکرد کلی و همچنین معیارهای فراخوان بهتر عمل کرد. این روش به ترتیب ۸۳/۸۱، ۸۱/۵۹ و ۸۲/۶۸ درصد در معیارهای ارزیابی دقت، یادآوری و اندازه گیری f1 به دست آورد. تا جایی که ما می دانیم، هیچ یک از مدل های موجود در هر سه معیار زمانی که در معیار DDIExtraction-2013 به کار گرفته شد، به بیش از ۸۰ درصد آستانه دست نیافتند. فراتر از بهبود دقت، ما از مکانیسم استخراج شدت برای DDI هایی استفاده کردیم که توسط مدل ما به عنوان DDI های مثبت (برهمکنش های واقعی) گزارش و طبقه بندی می شوند. از متن گزارش شده DDI ها در ادبیات، ما شدت هر برهمکنش را با استفاده از استراتژی تحلیل احساسات استخراج کردیم. ما شدت برهمکنش را با محاسبه قطبیت متن مورد استفاده برای گزارش یک برهمکنش بررسی کردیم. این مکانیسم به ما کمک کرد تا مجموعه داده جدیدی را در رابطه با وظیفه استخراج شدت DDI فرموله کنیم. در نتیجه، پیش بینی مکانیسم استخراج شدت با اعمال تحلیل های مختلف بر اساس تکلیف طبقه بندی DDI اعمال شده در DrugBank و MEDLINE به طور جداگانه ارزیابی و بررسی می شود.

۶. منابع و مراجع

1. Kohn, L.T.; Corrigan, J.M.; Donaldson, M.S. (Eds.) *To Err Is Human*; National Academies Press: Washington, DC, USA, 2000. [CrossRef]
2. Patanè, M.; Ciriaco, M.; Chimirri, S.; Ursini, F.; Naty, S.; Grembiale, R.D.; Gallelli, L.; De Sarro, G.; Russo, E. Interactions among Low Dose of Methotrexate and Drugs Used in the Treatment of Rheumatoid Arthritis. *Adv. Pharmacol. Sci.* 2013, 2013, 313858. [CrossRef] [PubMed]
3. Gallelli, L.; Staltari, O.; Palleria, C.; Di Mizio, G.; De Sarro, G.; Caroleo, B. A case of adverse drug reaction induced by dispensing error. *J. Forensic Leg. Med.* 2012, 19, 497–498. [CrossRef] [PubMed]
4. Leporini, C.; Patanè, M.; Saullo, F.; Rende, P.; Gallelli, L.; Di Paola, E.D.; Toscano, R.; Lucia, M.; Rossi, M.; De Sarro, G.; et al. A Comprehensive Safety Evaluation of Trabectedin and Drug–Drug Interactions of Trabectedin-Based Combinations. *BioDrugs* 2014, 28, 499–511. [CrossRef] [PubMed]
5. Staltari, O.; Leporini, C.; Caroleo, B.; Russo, E.; Siniscalchi, A.; De Sarro, G.; Gallelli, L. Drug-Drug Interactions: Antiretroviral Drugs and Recreational Drugs. *Recent Pat. CNS Drug Discov.* 2015, 9, 153–163. [CrossRef] [PubMed]
6. Palleria, C.; Di Paolo, A.; Giofrè, C.; Caglioti, C.; Leuzzi, G.; Siniscalchi, A.; De Sarro, G.; Gallelli, L. Pharmacokinetic drug-drug interaction and their implication in clinical management. *J. Res. Med. Sci. Off. J. Isfahan Univ. Med. Sci.* 2013, 18, 601–610.
7. Akram, J.; Kouzani, A.Z.; Mahmud, M.A.P. Using Adaptive Sensors for Optimised Target Coverage in Wireless Sensor Networks. *Sensors* 2022, 22, 1083. [CrossRef]
8. Leape, L.L.; Berwick, D.M. Five Years After to Err Is Human. *JAMA* 2005, 293, 2384. [CrossRef]
9. Lester, D. Theories of attempted suicide: Should they differ from theories of completed suicide? *Clin. Neuropsychiatry* 2009, 6, 188–191.
10. Tahir, A.; Akram, J.; Adil, M.; Ali, S.; Kouzani, A.Z.; Mahmud, M.A.P. Automatic Target Detection from Satellite Imagery Using Machine Learning. *Sensors* 2022, 22, 1147. [CrossRef]
11. Gallelli, L.; Nardi, M.; Prantera, T.; Barbera, S.; Raffaele, M.; Arminio, D.; Pirritano, D.; Colosimo, M.; Maselli, R.; Pelaia, G.; et al. Retrospective analysis of adverse drug reactions induced by gemcitabine treatment in patients with non-small cell lung cancer. *Pharmacol. Res.* 2004, 49, 259–263. [CrossRef] [PubMed]
12. de Vries, E.N.; Ramrattan, M.A.; Smorenburg, S.M.; Gouma, D.J.; Boermeester, M.A. The incidence and nature of in-hospital adverse events: A systematic review. *Qual. Saf. Health Care* 2008, 17, 216–223. [CrossRef] [PubMed]
13. Dollinger, C.; Schwiertz, V.; Sarfati, L.; Gourc-Berthod, C.; Guédât, M.-G.; Alloux, C.; Vantard, N.; Gauthier, N.; He, S.; Kiouris, E.; et al. SIMulation of Medication Error induced by Clinical Trial drug labeling: The SIMME-CT study. *Int. J. Qual. Health Care* 2016, 28, 311–315. [CrossRef] [PubMed]
14. Gallelli, L.; Maida, F.; Rende, P.; Russo, E.; Staltari, O.; Caroleo, B.; De Sarro, G. Recurrence of atrial fibrillation after switching from brand to generic atenolol. *J. Pharmacol. Pharmacother.* 2015, 6, 39. [CrossRef] [PubMed]
15. Qadir, Z.; Saeed, N.; Le, K. Towards 6G Internet of Things: Recent Advances, Use Cases, and Open Challenges. *arXiv* 2021, arXiv:2111.06596. [CrossRef]

16. De Vuono, A.; Scicchitano, F.; Palleria, C.; Russo, E.; De Sarro, G.; Gallelli, L. Lack of efficacy during the switch from brand to generic allopurinol. *J. Forensic Leg. Med.* 2013, 20, 540–542. [CrossRef]
17. Cai, R.; Liu, M.; Hu, Y.; Melton, B.L.; Matheny, M.E.; Xu, H.; Duan, L.; Waitman, L.R. Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports. *Artif. Intell. Med.* 2017, 76, 7–15. [CrossRef]
18. Qadir, Z.; Khan, S.I.; Khalaji, E.; Al-Turjman, F.; Mahmud, M.A.P.; Kouzani, A.Z.; Le, K. Predicting the energy output of hybrid PV–wind renewable energy system using feature selection technique for smart grids. *Energy Rep.* 2021, 7, 8465–8475. [CrossRef]
19. Aronson, J.K. Communicating information about drug interactions. *Br. J. Clin. Pharmacol.* 2007, 63, 637–639. [CrossRef]
20. Holzinger, A.; Dehmer, M.; Jurisica, I. Knowledge Discovery and interactive Data Mining in Bioinformatics-State-of-the-Art, future challenges and research directions. *BMC Bioinform.* 2014, 15 (Suppl. S6), I1. [CrossRef]
21. He, L.; Yang, Z.; Zhao, Z.; Lin, H.; Li, Y. Extracting Drug-Drug Interaction from the Biomedical Literature Using a Stacked Generalization-Based Approach. *PLoS ONE* 2013, 8, e65814. [CrossRef]
22. Herrero-Zazo, M.; Segura-Bedmar, I.; Martínez, P.; Declerck, T. The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions. *J. Biomed. Inform.* 2013, 46, 914–920. [CrossRef] [PubMed]
23. Khan, S.I.; Qadir, Z.; Nayak, S.R.; Budati, A.K.; Verma, K.D.; Prakash, D. UAVs path planning architecture for effective medical emergency response in future networks. *Phys. Commun.* 2021, 47, 101337. [CrossRef]
24. Björne, J.; Kaewphan, S.; Salakoski, T. UTurku: Drug Named Entity Recognition and Drug-Drug Interaction Extraction Using SVM Classification and Domain Knowledge. In *Proceedings of the *SEM 2013-2nd Joint Conference on Lexical and Computational Semantics*, Atlanta, GA, USA, 14 June 2013; Volume 2, pp. 651–659.
25. Duda, S.; Aliferis, C.; Miller, R.; Statnikov, A.; Johnson, K. Extracting drug-drug interaction articles from MEDLINE to improve the content of drug databases. *AMIA Annu. Symp. Proc. AMIA Symposium*. AMIA Symp. 2005, 2005, 216–220.
26. Guo, S.; Ramakrishnan, N. Mining linguistic cues for query expansion. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009*, Hong Kong, China, 2–6 November 2009; p. 335. [CrossRef]
27. Aggarwal, R.; Qadir, Z.; Khan, S.I.; Kouzani, A.Z.; Mahmud, M.A.P. A gabor filter-based protocol for automated image-based building detection. *Buildings* 2021, 11, 302. [CrossRef] *Big Data Cogn. Comput.* 2022, 6, 30 16 of 17
28. Kim, S.; Liu, H.; Yeganova, L.; Wilbur, W.J. Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach. *J. Biomed. Inform.* 2015, 55, 23–30. [CrossRef] [PubMed]
29. Akram, J.; Javed, A.; Khan, S.; Akram, A.; Ahmad, W. Swarm intelligence based localization in wireless sensor networks. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, New York, NY, USA, 22 April 2021; pp. 1906–1914. [CrossRef]
30. Bobi, T.; Fluck, J.; Hofmann-Apitius, M. SCAI: Extracting drug-drug interactions using a rich feature vector. In *Proceedings of the *SEM 2013-2nd Joint Conference on Lexical and Computational Semantics*, Atlanta, GA, USA, 13 June 2013; Volume 2, pp. 675–683. Available online: <https://aclanthology.org/S13-2111> (accessed on 14 December 2021).

31. Akram, J.; Tahir, A.; Akram, A.; Kouzani, A.Z.; Mahmud, M.A.P. Cloud- and Fog-Integrated Smart Grid Model for Efficient Resource Utilisation. *Sensors* 2021, 21, 7846. [CrossRef]
32. Iqbal, K.; Inam, H.; Qayyum, S. Promoting Customer Loyalty and Satisfaction in Financial Institutions through Technology Integration: The Roles of Service Quality, Awareness, and Perceptions. *Sustainability* 2021, 13, 12951. [CrossRef]
33. Zheng, W.; Lin, H.; Zhao, Z.; Xu, B.; Zhang, Y.; Yang, Z.; Wang, J. A graph kernel based on context vectors for extracting drug–drug interactions. *J. Biomed. Inform.* 2016, 61, 34–43. [CrossRef]
34. Shen, Y.; He, X.; Gao, J.; Deng, L.; Mesnil, G. Learning semantic representations using convolutional neural networks for web search. In *Proceedings of the 23rd International Conference on World Wide Web, New York, NY, USA, 7 April 2014*; pp. 373–374. [CrossRef]
35. Qadir, Z.; Munir, A.; Ashfaq, T.; Khan, M.A.; Le, K. A prototype of an energy-efficient MAGLEV train: A step towards cleaner train transport. *Clean. Eng. Technol.* 2021, 4, 100217. [CrossRef]
36. Yih, W.; He, X.; Meek, C. Semantic Parsing for Single-Relation Question Answering. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Baltimore, MD, USA, 26 June 2014*; Volume 2, pp. 643–648. [CrossRef]
37. Liu, S.; Tang, B.; Chen, Q.; Wang, X. Drug-Drug Interaction Extraction via Convolutional Neural Networks. *Comput. Math. Methods Med.* 2016, 2016, 103432. [CrossRef]
38. Liu, S.; Chen, K.; Chen, Q.; Tang, B. Dependency-based convolutional neural network for drug-drug interaction extraction. In *Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Shenzhen, China, 15–18 December 2016*; pp. 1074–1080. [CrossRef]
39. Huang, D.; Jiang, Z.; Zou, L.; Li, L. Drug–drug interaction extraction from biomedical literature using support vector machine and long short term memory networks. *Inf. Sci.* 2017, 415–416, 100–109. [CrossRef]
40. Quan, C.; Hua, L.; Sun, X.; Bai, W. Multichannel Convolutional Neural Network for Biological Relation Extraction. *BioMed Res. Int.* 2016, 2016, 2571–2581. [CrossRef] [PubMed]
41. Zhao, Z.; Yang, Z.; Luo, L.; Lin, H.; Wang, J. Drug drug interaction extraction from biomedical literature using syntax convolutional neural network. *Bioinformatics* 2016, 32, btw486. [CrossRef] [PubMed]
42. Zhang, D.; Wang, D. Relation Classification via Recurrent Neural Network. *arXiv* 2015, arXiv:1508.01006.
43. Khan, S.I.; Qadir, Z.; Kiani, Y.S.; Kouzani, A.Z.; Mahmud, M.A.P. Insights into the Mobility Pattern of Australians during COVID-19. *Sustainability* 2021, 13, 9611. [CrossRef]
44. Li, F.; Zhang, M.; Fu, G.; Ji, D. A neural joint model for entity and relation extraction from biomedical text. *BMC Bioinform.* 2017, 18, 198. [CrossRef]
45. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015-Conference Track Proceedings, San Diego, CA, USA, 1 January 2015*. Available online: <http://arxiv.org/abs/1409.0473> (accessed on 16 December 2021).
46. Santos, C.D.; Tan, M.; Xiang, B.; Zhou, B. Attentive Pooling Networks. *arXiv* 2016, arXiv:1602.03609.

47. Rocktäschel, T.; Grefenstette, E.; Hermann, K.M.; Kočiský, T.; Blunsom, P. Reasoning about Entailment with Neural Attention. In Proceedings of the 4th International Conference on Learning Representations, ICLR 2016-Conference Track Proceedings, San Juan, PR, USA, 22 September 2015. Available online: <http://arxiv.org/abs/1509.06664> (accessed on 12 December 2021).
48. Liaquat, M.U.; Rahman, A.; Qadir, Z.; Kouzani, A.Z.; Mahmud, M.A.P. Localization of Sound Sources: A Systematic Review. *Energies* 2021, 14, 3910. [CrossRef]
49. Liaquat, M.U.; Rahman, A.; Qadir, Z.; Kouzani, A.Z.; Mahmud, M.A.P. Sound Localization for Ad-Hoc Microphone Arrays. *Energies* 2021, 14, 3446. [CrossRef]
50. Niu, Y.; Zhu, X.; Li, J.; Hirst, G. Analysis of polarity information in medical text. *AMIA Annu. Symp. Proc.* 2005, 2005, 570–574.
51. Xia, L.; Gentile, A.L.; Munro, J.; Iria, J. Improving Patient Opinion Mining through Multi-step Classification. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5729, pp. 70–76. [CrossRef]
52. Romagnoli, K.M.; Nelson, S.D.; Hines, L.; Empey, P.; Boyce, R.D.; Hochheiser, H. Information needs for making clinical recommendations about potential drug-drug interactions: A synthesis of literature review and interviews. *BMC Med. Inform. Decis. Mak.* 2017, 17, 21. [CrossRef]
53. Khan, S.I.; Anum, N.; Qadir, Z.; Kouzani, A.Z.; Parvez Mahmud, M.A. Post-Flood Risk Management and Resilience Building Practices: A Case Study. *Appl. Sci.* 2021, 11, 4823. [CrossRef] *Big Data Cogn. Comput.* 2022, 6, 30 17 of 17
54. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; Dean, J. Distributed Representations of Words and Phrases and their Compositionality. *arXiv* 2013, arXiv:1310.4546.
55. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. In Proceedings of the 1st International Conference on Learning Representations, ICLR 2013-Workshop Track Proceedings, Scottsdale, AZ, USA, 17 January 2013. Available online: <http://arxiv.org/abs/1301.3781> (accessed on 2 January 2022).
56. Shaukat, M.A.; Shaukat, H.R.; Qadir, Z.; Kouzani, A.Z.; Mahmud, M.A.P. Cluster analysis and model comparison using smart meter data. *Sensors* 2021, 21, 3157. [CrossRef] [PubMed]
57. Khan, S.I.; Qadir, Z.; Kouzani, A.Z.; Mahmud, M.A.P. Insight into the Impact of COVID-19 on Australian Transportation Sector: An Economic and Community-Based Perspective. *Sustainability* 2021, 13, 1276. [CrossRef]
58. Wang, L.; Cao, Z.; de Melo, G.; Liu, Z. Relation Classification via Multi-Level Attention CNNs. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 13 August 2016; Volume 3, pp. 1298–1307. [CrossRef]
59. Muneeb, T.H.; Sahu, S.; Anand, A. Evaluating distributed word representations for capturing semantics of biomedical concepts. In Proceedings of the BioNLP 15, Beijing, China, 19 July 2015; pp. 158–163. [CrossRef]
60. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 2014, 15, 1929–1958.
61. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* 1997, 9, 1735–1780. [CrossRef]

62. Baccianella, S.; Esuli, A.; Sebastiani, F. SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010, Valletta, Malta, 9 May 2010; pp. 2200–2204.
63. Bokharaeian, B.; Díaz, A. NIL UCM: Extracting Drug-Drug interactions from text through combination of sequence and tree kernels. In Proceedings of the *SEM 2013-2nd Joint Conference on Lexical and Computational Semantics, Atlanta, GA, USA, 14 August 2013; Volume 2, pp. 644–650.
64. Rastegar-Mojarad, M.; Boyce, R.D.; Prasad, R. UWM-TRIADS: Classifying Drug-Drug Interactions with Two-Stage SVM and Post-Processing. In Proceedings of the *SEM 2013-2nd Joint Conference on Lexical and Computational Semantics, Atlanta, GA, USA, 12 June 2013; Volume 2, pp. 667–674.
65. Hailu, N.D.; Hunter, L.E.; Bretonnel Cohen, K. UColorado SOM: Extraction of Drug-Drug Interactions from BioMedical Text using Knowledge-rich and Knowledge-poor Features. In Proceedings of the *SEM 2013-2nd Joint Conference on Lexical and Computational Semantics, Atlanta, GA, USA, 15 June 2013; Volume 2, pp. 684–688.
66. Bui, Q.-C.; Sloot, P.M.A.; van Mulligen, E.M.; Kors, J.A. A novel feature-based approach to extract drug-drug interactions from biomedical text. *Bioinformatics* 2014, 30, 3365–3371. [CrossRef]
67. Raihani, A.; Laachfoubi, N. Extracting drug-drug interactions from biomedical text using a feature-based kernel approach. *J. Theor. Appl. Inf. Technol.* 2016, 92, 109–120.
68. Zhang, J.; Li, H.; Mo, D.; Chang, L. Mining Multispectral Aerial Images for Automatic Detection of Strategic Bridge Locations for Disaster Relief Missions. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Nature Switzerland AG: Cham, Switzerland, 2019; Volume 11607, pp. 189–200. [CrossRef]
69. Sun, X.; Dong, K.; Ma, L.; Sutcliffe, R.; He, F.; Chen, S.; Feng, J. Drug-Drug Interaction Extraction via Recurrent Hybrid Convolutional Neural Networks with an Improved Focal Loss. *Entropy* 2019, 21, 37. [CrossRef]
70. Ali Awan, A.; Khalid, U.; Maqsood, A. Revolutionizing Telemedicine by Instilling H.265. *Int. J. Image Graph. Signal Processing* 2017, 9, 20–27. [CrossRef]
71. Wu, H.; Xing, Y.; Ge, W.; Liu, X.; Zou, J.; Zhou, C.; Liao, J. Drug-drug interaction extraction via hybrid neural networks on biomedical literature. *J. Biomed. Inform.* 2020, 106, 103432. [CrossRef] [PubMed]
72. Park, C.; Park, J.; Park, S. AGCN: Attention-based graph convolutional networks for drug-drug interaction extraction. *Expert Syst. Appl.* 2020, 159, 113538. [CrossRef]

Investigating Big Data Management in Drug-Drug Interactions Based on Modern Deep Learning Approach for Smart Medicine and Healthcare

Mahdi Mirmasoumi^{*1}

1- MSc., Department of Accounting, Shamim Danesh Novin Institute of Higher Education, Ardabil, Iran

Abstract

Detecting and classifying drug-drug interactions (DDI) from available data is of great importance as reports indicate that DDI is one of the leading causes of hospital-acquired conditions. Also, for smart healthcare to avoid adverse drug interactions, having up-to-date knowledge about DDI is essential. This knowledge can be extracted using text processing techniques in the medical literature in the form of "big data". The medical literature is published in huge volumes, and it is impossible for health care professionals to collect all the peer-reviewed reports from this big data. To avoid this time-consuming method, information extraction (IE) and relation extraction (RE) techniques, which have been deeply studied in natural language processing (NLP). Current state-of-the-art implementations for extracting DDIs from biomedical texts use support vector machines (SVMs) or other machine learning methods that operate on manually defined features. Modern deep learning techniques have also been applied to automatically extract DDIs from scientific texts and have proven to be very promising for the advancement of DDI extraction tasks. We proposed a deep neural network-based method (SEV-DDI: Drug Interference Severity) with more integrated layers to achieve higher precision and accuracy. The ability to determine DDI severity would be very useful for more accurate and informed clinical decision support systems.

Keywords: Drug-drug interaction, Information extraction, Deep learning, Medicine, Smart healthcare