Journal of Information and Communication Technology

Volume 17, Issue 65, Special Issue on "Artificial Intelligence and data analytics", October 2025, pp 41-61

Improving Sentiment Analysis in Persian text based on combination of Stacked Auto-Encoder and Transformer-BiLSTM-CNN

Sina Dami[1*], MohammadAli Sanagooye Moharrer[2]

[1] Assistant Professor of Computer Engineering, West Tehran Branch, Islamic Azad University, Tehran, Iran
[2] MSc Graduate of Computer Engineering, West Tehran Branch, Islamic Azad University, Tehran, Iran

Received: 05 November 2024, Revised: 29 January 2025, Accepted: 30 April 2025
Paper type: Research

Abstract

The expansion of the internet and the increasing amount of user-generated textual opinions on various topics have made sentiment analysis a crucial tool for understanding public sentiment towards different subjects. These insights are invaluable for businesses, policymakers, and society as a whole, but manually analyzing such a volume of data is costly and impractical. This study leverages automated and deep learning approaches by combining a Stacked Autoencoder (SAE) for feature extraction and a Transformer-BiLSTM-CNN model for sentiment classification, specifically designed for the Persian language. ParsBert, the Persian version of BERT, was used for data preprocessing. This combined approach demonstrated improved performance in key evaluation metrics such as accuracy, precision, recall, and F1 score, outperforming comparative models like Transformer-BiLSTM-CNN, SAE-LSTM, and CNN. Results on datasets including user reviews from the Taghcheh and Digikala platforms and Persian tweets affirm the effectiveness of this hybrid model.

Keywords: Sentiment Analysis, Feature Extraction, Transformer, Stacked Auto-Encoder.

[*] Corresponding Author's email : dami@wtiau.ac.ir

دوفصلنامه
فناوری اطلاعات و ارتباطات ایران

سال هفدهم، شماره ۶۵، مهر ۱۴۰۴، شمارهٔ ویژهٔ «هوش مصنوعی و تحلیل داده»، صفحه ۴۱ الی ۶۱

بهبود تحلیل احساسات در متن فارسی براساس ترکیبی از رمزگذار خودکار پشته‌ای و Transformer-BiLSTM-CNN

سینا دامی[1*]، محمدعلی ثناگوی محرر[2]

[1] استادیار گروه مهندسی کامپیوتر، واحد تهران غرب، دانشگاه آزاد اسلامی، تهران، ایران

[2] دانش‌آموخته ارشد مهندسی کامپیوتر، واحد تهران غرب، دانشگاه آزاد اسلامی، تهران، ایران

تاریخ دریافت: ۱۴۰۳/۰۸/۱۵ تاریخ بازبینی: ۱۴۰۳/۱۱/۱۰ تاریخ پذیرش: ۱۴۰۴/۰۲/۱۰

نوع مقاله: پژوهشی

چکیده

با گسترش اینترنت و افزایش ثبت نظرات متنی کاربران درباره رویدادهای مختلف، تحلیل احساسات متنی اهمیت بالایی یافته است، زیرا می‌تواند احساسات مثبت، منفی یا خنثی کاربران را نسبت به موضوعات گوناگون نشان دهد و به شرکت‌ها، سیاست‌گذاران و جامعه، اطلاعات ارزشمندی ارائه دهد. اما تحلیل دستی این حجم از داده‌ها بسیار زمان‌بر و پرهزینه است؛ لذا پژوهشگران به روش‌های خودکار و یادگیری عمیق روی آورده‌اند. این پژوهش با ترکیب رمزگذار خودکار پشته‌ای (SAE) برای استخراج ویژگی‌ها و مدل ترکیبی Transformer-BiLSTM-CNN برای طبقه‌بندی، به بهبود تحلیل احساسات متنی در زبان فارسی پرداخته است. برای پیش‌پردازش داده‌ها نیز از مدل ParsBert، نسخه فارسی BERT، استفاده شد. این ترکیب در برخی معیارهای ارزیابی مانند دقت، صحت، یادآوری و امتیاز F1، نسبت به مدل‌های مشابه همچون Transformer-BiLSTM-CNN، SAE-LSTM و CNN عملکرد بهتری داشت و نتایج مطلوبی در مجموعه‌داده‌های نظرات کاربران طاقچه و دیجی‌کالا و توئیت‌های فارسی به دست آمد.

کلیدواژگان: تحلیل احساسات، استخراج ویژگی، تبدیل کننده، رمزگذار خودکار پشته‌ای.

* رایانامهٔ نویسنده مسؤول: dami@wtiau.ac.ir

۱- مقدمه

با توسعه‌ی روزافزون شبکه‌های مجازی و تجارت الکترونیک، استفاده‌ی مردم از خدمات مبتنی‌بر وب افزایش پیدا کرده است. کاربران از فروشگاه‌های اینترنتی خرید می‌کنند و درباره‌ی محصولات، نظر می‌نویسند؛ این نظرات بر روی سایر خریداران نیز تاثیر می‌گذارد و به این ترتیب مردم برای خرید یک محصول، خود را محدود به نظرات خانواده و اطرافیان خود نمی‌کنند. مردم علاقه دارند درباره‌ی موضوعات مختلف صحبت کنند و احساسات خود را در مسائلی مانند سیاست بیان کنند. فضای مجازی باعث حضور اشخاص تاثیرگذار از حوزه‌های مختلف اعم از صاحبان کسب و کارها تا سیاستمداران شده که همین موجب می‌شود مردم نظرات مثبت، منفی و یا خنثی خود را درباره‌ی آن‌ها ثبت کنند؛ نظراتی که می‌تواند ناشی از عملکرد مثبت یا منفی آن‌ها باشد. تحلیل احساسات به دلیل کاربردهای گسترده‌ی خود، توجه زیادی از حوزه‌های صنعتی و دانشگاهی به خود جلب کرده است. تحلیل احساسات به طور خلاصه شامل تحلیل محاسباتی متن در پردازش زبان طبیعی[1] (NLP) برای شناسایی، استخراج و طبقه‌بندی اطلاعات ذهنی از متن بدون ساختار است [۱].

عقیده کاوی کاربران می‌تواند کاربردهای فراوانی داشته باشد؛ به عنوان مثال فروشگاه‌های اینترنتی با تحلیل این اطلاعات می‌توانند تقاضاهای روز و آینده‌ی کاربران را پیشبینی کنند و محصولات را مطابق با آن تامین کنند، تولیدکننده‌ها و ارائه‌دهندگان خدمات، با استفاده از احساسات مثبت و منفی کاربران درباره‌ی محصولات و خدمات خود، می‌توانند نقاط ضعف و قوت خود را شناسایی کنند و کالا و خدمات ضعیف خود را بهبود ببخشند یا حذف کنند، یا تولید کالا یا ارائه خدمات محبوب خود را افزایش دهند یا از احساسات مثبت و منفی کاربران نسبت به موضوعات مختلف، برای محصولات و خدمات جدید خود ایده بگیرند. دولت‌ها می‌توانند از رضایت یا عدم رضایت مردم از موضوع خاصی مطلع شده وقوع بحران‌های امنیت ملی را پیشبینی و از آن جلوگیری کنند. اینکه بدانیم تظاهرات هنگ کنگ ۲۰۲۰، لندن [۲] و بهار عربی ۲۰۱۱ [۳] به وسیله‌ی شبکه‌های اجتماعی اینترنتی سازماندهی شد؛ ضرورت پایش، جمع‌آوری و تحلیل احساسات کاربران این شبکه‌ها را دوچندان می‌کند. از مهم‌ترین کاربردهای تحلیل احساسات همچنین می‌توان به پیش‌بینی یا تاثیرگذاری بر انتخابات و سود سهام اشاره کرد [۴].

تحلیل احساسات کاربران از متن‌های تولیدشده در فضای مجازی به یکی از ابزارهای کلیدی برای درک رفتار و ترجیحات آن‌ها تبدیل شده است. در این میان، تحلیل احساسات متنی به زبان فارسی به دلیل پیچیدگی‌های زبانی و کمبود منابع مناسب، با چالش‌های خاصی مواجه است. بهبود عملکرد مدل‌های تحلیل احساسات در این حوزه نه تنها می‌تواند به کسب‌وکارها و دولت در تصمیم‌گیری‌های استراتژیک کمک کند، بلکه به توسعه‌ی ابزارهای پردازش زبان طبیعی برای زبان‌های کم‌منبع نیز یاری می‌رساند.

هدف اصلی این پژوهش، ارائه یک مدل ترکیبی جدید برای بهبود دقت و کارایی تحلیل احساسات در متون فارسی است. با توجه به چالش‌های خاص زبان فارسی مانند پیچیدگی‌های صرفی و نحوی، و همچنین کمبود منابع پردازش زبان طبیعی در این زبان، نیاز به مدل‌های تخصصی و کارآمد احساس می‌شود. ما قصد داریم با ترکیب هوشمندانه مدل‌های یادگیری عمیق که در تحقیقات پیشین عملکرد قابل توجهی داشته‌اند، یک معماری جدید ارائه دهیم که بتواند عملکرد بهتری نسبت به مدل‌های موجود داشته باشد.

فرضیه اصلی ما در این پژوهش این است که ترکیب مدل‌های بهینه‌شده می‌تواند عملکرد تحلیل احساسات متنی را بهبود بخشد. این فرض بر این اساس شکل گرفته که هر یک از مدل‌های ParsBERT، رمزگذار خودکار پشته‌ای و Transformer-BiLSTM-CNN شبکه عصبی کانولوشنال[2] (CNN) در بخش‌های مختلف پردازش متن (پیش‌پردازش، استخراج ویژگی و طبقه‌بندی) نقاط قوت منحصر به فردی دارند. ParsBert به دلیل توانایی آن در پردازش متن‌های فارسی و استخراج ویژگی‌های معنایی عمیق استفاده شده است. همچنین، استفاده از رمزگذار خودکار پشته‌ای[3] (SAE) به دلیل قابلیت آن در کاهش ابعاد داده‌ها و استخراج ویژگی‌های غیرخطی صورت گرفته است. ترکیب مدل‌های Transformer-BiLSTM-CNN نیز به این دلیل انتخاب شده است که Transformer توانایی پردازش متن‌های طولانی و وابستگی‌های بلندمدت را دارد، شبکه حافظه بلند و کوتاه مدت دو جهته[4] (BiLSTM) برای مدل‌سازی ترتیب کلمات و ویژگی‌های متوالی مناسب است؛ و CNN نیز برای پردازش و تشخیص ویژگی‌های محلی و الگوهای موثر در تحلیل احساسات به‌کار می‌رود. این ترکیب به ما امکان می‌دهد تا از نقاط قوت هر مدل بهره‌برداری کرده و عملکرد کلی تحلیل احساسات را بهبود بخشیم.

نوآوری اصلی این پژوهش در ترکیب هوشمندانه سه مدل قدرتمند

[1] Natural language processing
[2] Convolutional Neural Network
[3] Stacked Auto Encoder
[4] Bidirectional Long Short-Term Memory Network

یادگیری عمیق است که هر کدام در حوزه خود عملکرد برجسته‌ای داشته‌اند. این ترکیب نه تنها باعث بهره‌برداری از نقاط قوت هر مدل می‌شود، بلکه با کنترل پیچیدگی محاسباتی، یک راه‌حل کارآمد برای تحلیل احساسات در متون فارسی ارائه می‌دهد.

ما در این پژوهش همچنین تلاش می‌کنیم تا به سوالات زیر پاسخ دهیم:

۱. چگونه با ترکیب مدل‌های یادگیری عمیق، عملکرد وظیفه تحلیل احساسات را در متن‌های فارسی بهبود دهیم؟

۲. چگونه می‌توان با بکارگیری مدل‌های بسیار عمیق باعث بهبود استخراج ویژگی‌ها در وظیفه تحلیل احساسات شد؟

در ادامه، نوآوری‌ها و دستاوردهای اصلی این تحقیق توضیح داده می‌شوند:

۱) در این پژوهش، برای اولین بار ترکیبی از مدل‌های Transformer-BiLSTM-CNN با استفاده از SAE ارائه شده است. این ترکیب، مزایای مدل‌های مبتنی بر توجه برای پردازش و تشخیص روابط معنایی عمیق، توانایی BiLSTM در یادگیری وابستگی‌های زمانی بلندمدت، و قابلیت CNN در پردازش و تشخیص ویژگی‌های محلی را با یکدیگر ادغام کرده است. همچنین، استفاده از SAE به بهبود بازنمایی داده‌ها و کاهش ابعاد ویژگی‌ها کمک کرده است.

۲) این پژوهش از ParsBERT به عنوان ابزار اصلی پیش‌پردازش و ایجاد بردارهای معنایی استفاده کرده است. این مدل، به دلیل تمرکز بر ویژگی‌های زبانی و ساختاری خاص زبان فارسی، امکان تحلیل دقیق‌تر احساسات را فراهم کرده است. همچنین، تکنیک‌های پیش‌پردازش خاصی نظیر نرمال‌سازی و حذف نویز به منظور بهبود کیفیت داده‌ها اعمال شده است.

۳) یکی از جنبه‌های نوآورانه این پژوهش، ارزیابی مدل در شرایط داده‌های کم‌منبع است. این پژوهش نشان داده است که ترکیب مدل‌های پیشنهادی قادر است با داده‌های محدود نیز عملکرد قابل قبولی داشته باشد و این موضوع به‌ویژه در تحلیل متون زبان فارسی اهمیت بالایی دارد.

۴) توسعه معماری برای کاربردهای چندزبانه: اگرچه تمرکز اصلی این پژوهش بر تحلیل متون فارسی است، ساختار پیشنهادی با تنظیمات مناسب می‌تواند به راحتی برای زبان‌های دیگر نیز به کار رود. این ویژگی، پتانسیل گسترش روش در سایر حوزه‌ها و زبان‌ها را فراهم می‌کند.

۵) کاهش نیاز به منابع محاسباتی: یکی از دستاوردهای کلیدی این پژوهش، استفاده از SAE برای کاهش ابعاد داده و بهبود بازنمایی ویژگی‌ها است که به طور مستقیم نیاز به منابع محاسباتی را کاهش داده و سرعت پردازش را افزایش داده است. این موضوع، امکان استفاده از مدل پیشنهادی در محیط‌های محدود به منابع را فراهم می‌کند.

ادامه‌ی ساختار این مقاله به‌شرح زیر است: در بخش دوم، مفاهیم اولیه و مبانی نظری مرتبط با تحلیل احساسات و روش‌های یادگیری عمیق مورد بررسی قرار می‌گیرد. در بخش سوم، مروری بر پژوهش‌های مرتبط انجام می‌شود. در بخش چهارم، روش تحقیق و مدل پیشنهادی به‌طور مفصل شرح داده می‌شود. در بخش پنجم، ارزیابی مدل پیشنهادی با استفاده از معیارهای استاندارد و مجموعه داده‌های مناسب ارائه می‌شود و در نهایت، در بخش ششم نتیجه‌گیری و پیشنهادات برای پژوهش‌های آینده ارائه خواهد شد.

۲- مفاهیم اولیه

در این قسمت مفاهیم اولیه شامل تعریف تحلیل احساسات متنی و یادگیری عمیق ارائه شده است و در ادامه شبکه‌های عصبی مورد استفاده در این مقاله معرفی می‌شود.

۲-۱- تحلیل احساسات متنی

محبوبیت اینترنت و سهولت دسترسی آن برای همه‌ی مردم، محتوای فراوانی را در فضای آنلاین تولید کرده است. مردم احساسات و نظرات خود را در شبکه‌های اجتماعی از طریق متن، تصویر و ویدئو بیان می‌کنند. همچنین نظرات خود را در مورد محصولات و خدمات خاص از طریق نظرات تعاملی در وب سایت‌های تجاری و بستر شبکه‌های اجتماعی بروز می‌دهند. محتوای متنی تولید شده در این بستر، بینش ارزشمندی در مورد نظر و احساسات نظردهنده در مورد موضوع مورد بحث، به دست می‌دهد. این نظرات به وسیله‌ی تکنیک‌های پردازش زبان طبیعی بهتر درک می‌شود و مورد استفاده قرار می‌گیرند. احساس معادل نشان دادن نظر و ارزیابی و نگرش است [۵] و تحلیل احساسات، فرآیند شناسایی، استخراج و طبقه‌بندی اطلاعاتی با مفاهیم خاص از محتوای متنی بدون ساختار است که به آن عقیده‌کاوی نیز گفته می‌شود و لحن عاطفی یا احساسات منتقل شده در داده‌های متنی را تعریف می‌کند [۶].

هدف از طبقه‌بندی احساسات، استخراج احساسات و نظرات از متن‌ها است. وظیفه‌ی تحلیل احساسات یک مسئله طبقه‌بندی متن محسوب می‌شود؛ زیرا متن را به دو دسته‌ی احساس مثبت و احساس منفی دسته‌بندی می‌کند. تشخیص طعنه، ذهنیت و سازماندهی متن از جمله مواردی هستند که این وظیفه را دشوار می‌کنند [۷].

۲-۲- یادگیری عمیق

در سال‌های اخیر الگوریتم‌های یادگیری عمیق به نتایج قابل توجهی در حوزه پردازش زبان طبیعی دست یافتند. اصطلاح "یادگیری عمیق" به شبکه‌های عصبی با لایه‌های متعدد پرسپترون الهام گرفته از مغز ما اشاره دارد و داده‌ها را در لایه‌های متعدد و متوالی نشان می‌دهند؛ تعداد لایه‌ها معیار مهمی برای نمایش عمق شبکه است. با این معماری می‌توان مدل‌های پیچیده‌تری را بر روی مجموعه داده‌های بسیار بزرگ‌تری آموزش داد و نتایج پیشرفته‌تری را در بسیاری از حوزه‌های کاربردی، از بینایی رایانه و تشخیص گفتار گرفته تا NLP، تولید کرد [۷]. همچنین این توانایی را دارند که ویژگی‌های نحوی جملات را به طور خودکار بدون تکنیک‌های استخراج ویژگی اضافی ذخیره کنند؛ بدون تکنیک‌هایی که منابع و زمان بیشتری را مصرف می‌کنند. به همین دلیل است که مدل‌های یادگیری عمیق توجه محققان پردازش زبان طبیعی را برای طبقه‌بندی احساسات به خود جلب کرده است [۱]. مجموعه داده‌های بزرگ در آموزش، پیشرفت در منابع سخت‌افزاری برای پردازش داده‌های بزرگ و بهبود در مدل‌های یادگیری عمیق باعث افزایش اثربخشی مدل‌های یادگیری عمیق می‌شود [۸].

۲-۲-۱- BiLSTM

شبکه عصبی BiLSTM یکی از الگوریتم‌های شبکه عصبی بازگشتی، برای بهبود LSTM است و کار مدل سازی متوالی[1] را بهتر از LSTM حل می‌کند. اطلاعات در LSTM، از عقب به جلو حرکت می‌کنند؛ اما در BiLSTM، اطلاعات در هر دو جهت حرکت می‌کنند؛ هم از عقب به جلو و هم از جلو به عقب، با استفاده از دو حالت پنهان در حرکت هستند. با این کار زمینه را بهتر یاد می‌گیرد و داده‌های ورودی اطلاعات گذشته و آینده حفظ می‌شود [۱].

۲-۲-۲- CNN

CNN یکی از مهمترین معماری‌های یادگیری عمیق است و یک مدل شبکه عصبی پیش‌خور[2] چند لایه است و توانایی محاسبات موازی دارد [۷]. شبکه‌هایی که همه گره‌ها را می‌توان در لایه‌های متوالی سازمان‌دهی کرد، به طوری که هر گره ورودی را فقط از گره لایه‌های قبلی دریافت کند، شبکه عصبی پیش‌خور نام دارند [۹]. CNN، ویژگی‌ها را بر اساس اصل فضایی[3] استخراج می‌کند. شبکه با این اصل می‌تواند الگوهای فضایی را در داده‌های ورودی تشخیص دهد و بهتر یاد بگیرد [۸]. CNN از سه لایه تشکیل شده است: لایه کانولوشن[4]، لایه ادغام[5] و لایه کاملاً متصل[6].

۲-۲-۳- تبدیل کننده‌ها

این مدل شامل بلوک‌های رمزگذار و رمزگشا با یک تابع فعال‌سازی سافت مکس[7] برای نرمال سازی احتمالات خروجی شبکه عصبی است. ورودی مدل یک داده متوالی است. بلوک‌های رمزگذار متشکل از توجه چند سر[8] و یک شبکه پیش‌خور است. لایه‌های توجه چندسر، بردارهای توجه را برای هر ورودی محاسبه می‌کنند. رمزگشا دارای رمزگذار موقعیتی و لایه‌های توجه چندسر پوشانده شده است که مشابه بلوک رمزگذار عمل می‌کنند. مدل تبدیل‌کننده، که در ابتدا برای ترجمه ماشینی طراحی شده بود، برای مدل‌سازی زبان استفاده می‌شود و آن را برای سایر وظایف NLP مانند طبقه‌بندی متن، خلاصه‌سازی اسناد و پاسخ‌گویی به سؤال، قابل استفاده می‌کند [۱۰].

۲-۲-۴- نمایش رمزگذار دو جهته از تبدیل کننده‌ها[9] (BERT)

BERT از رمزگذارهای موجود در یک تبدیل‌کننده به عنوان زیرساختی برای مدل‌های پیش‌آموزشی استفاده می‌کند و برای وظایف NLP مانند تحلیل احساسات، پاسخ به سؤالات، خلاصه‌سازی متن و... کاربرد دارد. مزایای BERT شامل توانایی آن در مدیریت استخراج اطلاعات متنی به دلیل قابلیت دو جهته آن است. سریع‌تر تمرین می‌کند و در طیف گسترده‌ای از برنامه‌های مدل‌سازی زبان استفاده شده است. با این حال، معایبی نیز دارد از جمله اینکه به طبقه‌بندی‌های تک زبانه محدود می‌شود، طول جملات ورودی نیز آن را محدود می‌کند، مشکل استنتاج عملی دارد [۱۰].

[1] task of sequential modeling
[2] feed-forward
[3] spatial principle
[4] convolution
[5] pooling
[6] fully connected
[7] Softmax
[8] multi-head attention
[9] Bidirectional encoder representations from Transformers

۲-۲-۵- رمزگذارهای خودکار

رمزگذار خودکار یک بلوک ساختمانی اساسی است که می‌تواند به صورت سلسله مراتبی برای ایجاد مدل‌های عمیق استفاده شود. آن‌ها ویژگی‌های سطح بالا را سازماندهی، فشرده و استخراج می‌کنند و امکان یادگیری بدون نظارت و استخراج ویژگی‌های غیرخطی را فراهم می‌کنند. رمزگذارهای خودکار دارای یک رمزگذار و یک رمزگشا هستند. رمزگذارهای خودکار شبکه‌های عصبی پیشخور هستند که به اطلاعات اجازه می‌دهند در یک جهت جریان پیدا کنند [۱۱].

SAE -۶-۲-۲

رمزگذار خودکار سنتی معمولاً از یک رمزگذار تک لایه استفاده می‌کند که استخراج ویژگی‌های عمیق را به چالش می‌کشد. برای تقویت استخراج ویژگی، یک استراتژی موثر، تعمیق ساختار شبکه عصبی است. با بکارگیری یک رویکرد یادگیری لایه‌ای، می‌توان چندین رمزگذار خودکار پایه را در کنار هم قرار داد تا یک SAE را تشکیل دهند که امکان استخراج ویژگی‌های داده پیچیده را فراهم می‌کند [۱۱]. SAE از دو رمزگذار و رمزگشا استفاده می‌شود که ویژگی‌های واقعی را به طور موثر کاهش می‌دهد؛ به طوری که طبقه‌بندی و انجام هر گونه عملیات بر روی داده‌های متنی آسان می‌شود [۱۲].

۳- مرور ادبیات و سوابق مربوطه

فضلی واحید و همکاران [۱۲]، مدلی ترکیبی و نوآورانه برای تحلیل احساسات ارائه کردند که از ترکیب دو روش قدرتمند SAE و LSTM بهره می‌برد. هدف اصلی این پژوهش، غلبه بر مشکلات مرتبط با استخراج ناکافی ویژگی‌های اطلاعاتی توسط LSTM و محدودیت‌های لایه سافت مکس در SAE برای طبقه‌بندی دقیق‌تر ویژگی‌ها بوده است. مدل پیشنهادی آن‌ها شامل استفاده از SAE با دو لایه رمزگذار و دو لایه رمزگشا است که برای استخراج ویژگی‌های اطلاعاتی و کاهش ابعاد داده‌ها طراحی شده است. پس از این مرحله، LSTM به کار گرفته می‌شود که به دلیل توانایی آن در یادگیری وابستگی‌های بلندمدت در متن، فرآیند طبقه‌بندی احساسات را براساس ویژگی‌های استخراج شده انجام می‌دهد. پیش‌پردازش داده‌ها در این مدل شامل حذف علائم نگارشی، کلمات توقف و توکن‌سازی متن است. نتایج به‌دست‌آمده نشان داد که مدل ترکیبی پیشنهادی دقت[1] ۸۷ درصدی داشته و نسبت به مدل‌های سنتی مانند LSTM با دقت ۸۲,۲ درصد و CNN با دقت ۴۵,۵ درصد عملکرد بهتری از خود نشان داده است. این برتری نه تنها در بهبود دقت طبقه‌بندی احساسات مشهود بود، بلکه در کاهش زمان اجرای مدل نیز تأثیرگذار بود. علاوه بر این، مدل در شرایط مختلفی از تقسیم داده‌های آموزشی و تستی نیز عملکرد مطلوبی داشت که نشان‌دهنده پایداری آن است. با این حال، محدودیت‌هایی در این پژوهش وجود داشت. به‌عنوان مثال، مدل تنها برای طبقه‌بندی دو کلاسه (مثبت و منفی) طراحی شده و قابلیت تحلیل احساسات چندکلاسه را ندارد. همچنین استفاده از مکانیزم‌های توجه برای بهبود تحلیل احساسات در آن بررسی نشده است.

مریگانک شوکلا و آخیل کومار [۶]، در پژوهشی جامع به مقایسه و ارزیابی مدل‌های مختلف شبکه‌های عصبی عمیق برای تحلیل احساسات پرداختند. هدف اصلی این پژوهش، شناسایی بهترین ترکیب از مدل‌های ترکیبی نظیر CNN، BiLSTM و Transformer همراه با انواع تعبیه‌های کلمات مانند Word2Vec، GloVe و FastText برای دستیابی به بالاترین دقت در تحلیل احساسات بود. در این مطالعه، هر یک از مدل‌ها نقش خاصی ایفا کردند. شبکه‌های CNN برای استخراج ویژگی‌های محلی از متن، BiLSTM برای یادگیری وابستگی‌های بلندمدت و اطلاعات دنباله‌ای، و Transformer برای بهره‌گیری از مکانیزم توجه و بهبود درک معنایی کلمات در متن به کار گرفته شدند. نتایج این پژوهش نشان داد که مدل Transformer-BiLSTM-CNN در ترکیب با تعبیه Word2Vec بهترین عملکرد را ارائه داده و به دقت ۸۹,۰۴٪ دست یافته است. این برتری به ویژه زمانی که از ترکیب چندین تعبیه کلمات به عنوان کانال ورودی استفاده شد، آشکارتر بود. همچنین، افزودن لایه‌های Transformer به مدل‌های CNN و BiLSTM دقت را حدود ۲ تا ۳ درصد افزایش داد. با وجود این موفقیت‌ها، پژوهش با محدودیت‌هایی نیز مواجه بود. برای نمونه، مدل‌های پیشنهادی در مقایسه با مدل‌های پیش‌آموزش‌دیده‌ای چون BERT و ELMo دقت کمتری داشتند و آموزش مدل‌های ترکیبی با لایه‌های Transformer نیاز به منابع محاسباتی بیشتری داشت. علاوه بر این، استفاده از یک مجموعه داده محدود SST-2 می‌توانست تعمیم‌پذیری نتایج را کاهش دهد. پژوهش دیگری که به بررسی جامع مدل‌های تحلیل احساسات پرداخته، بر نقش ساختار مدل‌ها و نوع ورودی داده‌ها در عملکرد این مدل‌ها تمرکز داشته است. در این مطالعه، هشت مدل مختلف شامل نسخه‌های ساده و عمیق CNN و مدل‌های شبکه عصبی بازگشتی[2] (RNN) نظیر LSTM و

[1] Accuracy

[2] recurrent neural network

واحد بازگشتی دروازه‌دار[1] (GRU) با ورودی‌های سطح کلمه و کاراکتر ارزیابی شده‌اند. نتایج نشان داد که مدل‌های RNN با ورودی‌های سطح کلمه عملکرد بهتری نسبت به CNN داشتند، در حالی که CNN با ورودی سطح کاراکتر در برخی موارد، به ویژه در ساختارهای عمیق، نتایج مطلوب‌تری ارائه دادند. این پژوهش نیز محدودیت‌هایی مانند عدم استفاده از بردارهای پیش‌آموزش‌دیده و بررسی ترکیب CNN و RNN را داشت.

سکرین تام و همکاران [۱]، در پژوهش خود مدل ترکیبی نوینی به نام ConvBiLSTM را برای طبقه‌بندی احساسات معرفی کرده‌اند که از ادغام CNN و BiLSTM بهره می‌برد. هدف اصلی این تحقیق، بهبود دقت و عملکرد در تحلیل احساسات متون، به‌ویژه در تحلیل توییت‌ها، بوده است. این مدل با استفاده از توانایی CNN در استخراج ویژگی‌های محلی و قابلیت BiLSTM در یادگیری وابستگی‌های بلند مدت متن، تلاش کرده است تا محدودیت‌های روش‌های پیشین را کاهش دهد. نتایج به‌دست آمده نشان داد که مدل ConvBiLSTM توانسته است در مقایسه با مدل‌های پایه نظیر CNN و BiLSTM، عملکرد بهتری ارائه دهد. این مدل در مجموعه داده توییت‌ها به دقت ۹۴,۱۳٪ و در مجموعه داده SST-2 به دقت ۹۱,۱۳٪ دست یافت، در حالی که مدل CNN به ترتیب دقت ۹۱,۸۹٪ و ۸۹,۱۸٪ و مدل BiLSTM دقت ۹۱,۵۲٪ و ۸۹,۴۲٪ را به ثبت رسانده بودند. با وجود موفقیت‌های این مدل، پژوهش با محدودیت‌هایی نیز مواجه بود. از جمله این محدودیت‌ها می‌توان به نیاز به منابع محاسباتی بالا برای داده‌های حجیم و عدم استفاده از مکانیزم‌های توجه اشاره کرد.

شفیق و همکاران [۱۳]، در مطالعه‌ای به تحلیل احساسات مبتنی بر جنبه[2] (ABSA) پرداختند که هدف اصلی آن شناسایی و تحلیل احساسات مرتبط با جنبه‌های خاص متن در زبان عربی بود. این تحقیق به‌طور ویژه به بررسی این سوال پرداخت که چگونه می‌توان با ترکیب روش‌های انتها به انتها[3] (E2E)، عملکرد مدل‌های ABSA را بهبود بخشید. برای دستیابی به این هدف، آن‌ها از ترکیب AraBERT، یک مدل پیش‌آموخته برای زبان عربی، با فیلدهای تصادفی شرطی[4] (CRF) استفاده کردند. این رویکرد با ادغام وظایف استخراج اصطلاحات جنبه[5] (ATE) و طبقه‌بندی احساسات مربوطه[6] (ASC) در یک مدل واحد، توانست ارتباط میان این دو زیروظیفه را تقویت کند. همچنین، برای ارزیابی و جلوگیری از بیش‌برازش[7]، از تکنیک اعتبار سنجی متقاطع[8] k-fold بهره بردند. نتایج نشان داد که مدل پیشنهادی توانست به دقت ۹۵,۱۱٪ در وظیفه E2E-ABSA دست یابد و به ترتیب دقت‌های ۹۷,۷۸٪ و ۹۸,۳۴٪ را در وظایف ATE و ASC کسب کند. این دستاورد، بهبود قابل‌توجهی نسبت به مدل‌های پیشین داشت که این دو وظیفه را جداگانه انجام می‌دادند. با این حال، این پژوهش با محدودیت‌هایی مواجه بود. نخست، محدودیت داده‌های موجود برای زبان عربی عملکرد مدل را تحت تأثیر قرار داد. دوم، مدل‌های انتها به انتها در تحلیل جملات پیچیده و طولانی کارایی کمتری داشتند. همچنین، آموزش مدل‌های پیش‌آموخته‌ای چون AraBERT نیاز به منابع محاسباتی بالایی داشت. از دیگر محدودیت‌های تحقیقاتی قابل‌توجه می‌توان به عدم توسعه مدل‌هایی اشاره کرد که بتوانند چندین زیروظیفه ABSA را به‌طور همزمان انجام دهند و در داده‌های کم‌منبع نیز عملکرد مناسبی داشته باشند.

محمد الجبرین و همکاران [۱۴]، در پژوهشی به تحلیل احساسات کاربران در شبکه‌های اجتماعی، به‌ویژه توییتر، پرداختند. هدف اصلی این تحقیق طبقه‌بندی خودکار احساسات کاربران نسبت به فناوری ChatGPT به سه دسته مثبت، منفی و خنثی بود. این مطالعه با ارائه یک مدل ترکیبی به نام MFOHDL-SA، از الگوریتم بهینه‌سازی پروانه شعله[9] (MFO) و یادگیری عمیق ترکیبی[10] بهره گرفت تا دقت و کارایی تحلیل احساسات را بهبود بخشد. برای دستیابی به این هدف، یک فرآیند جامع پیش‌پردازش داده‌ها طراحی شد که شامل حذف ریتوییت‌ها، URLها، علائم نگارشی، تبدیل ایموجی‌ها به کلمات، توکن‌سازی، نرمال‌سازی، حذف کلمات توقف و انجام Stemming و Lemmatization بود. سپس، از مدل TF-IDF برای استخراج ویژگی‌ها و بردارسازی داده‌ها استفاده شد. مدل ترکیبی CNN-LSTM برای طبقه‌بندی احساسات به کار گرفته شد و الگوریتم MFO برای تنظیم هایپرپارامترهای این مدل به منظور بهینه‌سازی عملکرد آن استفاده گردید. نتایج این پژوهش نشان داد که مدل MFOHDL-SA به دقت کلی ۹۵,۰۹٪ در طبقه‌بندی احساسات دست یافت. در مرحله تست، دقت مدل برای کلاس‌های مثبت، منفی و خنثی به ترتیب ۹۶,۸۵٪، ۹۴,۶۷٪ و ۹۳,۷۳٪ بود. این عملکرد به‌طور قابل‌توجهی بهتر از مدل‌های پایه‌ای مانند جنگل

[1] Gated Recurrent Unit
[2] Aspect-Based Sentiment Analysis
[3] End-to-End
[4] Conditional Random Fields
[5] Aspect Term Extraction
[6] Aspect Sentiment Classification
[7] Overfitting
[8] Cross validation
[9] Moth Flame Optimization
[10] Hybrid Deep Learning

تصادفی، درخت تصمیم، ماشین بردار پشتیبانی[1] (SVM)، XGBoost، CNN و ماشین یادگیری افراطی[2] (ELM) بود. با وجود این موفقیت‌ها، مطالعه با محدودیت‌هایی نیز مواجه بود. از جمله، حجم داده‌ها محدود به ۶۰،۰۰۰ نمونه بود که ممکن است برای آموزش مدل‌های پیچیده کافی نباشد. همچنین، مدل‌های ترکیبی مانند CNN-LSTM منابع محاسباتی زیادی مصرف می‌کنند و زمان آموزش طولانی‌تری دارند. علاوه بر این، الگوریتمMFO در مواجهه با داده‌های بسیار بزرگ یا پیچیده ممکن است در تنظیم بهینه‌های هایپرپارامتری، کند عمل کند.

ایزابل کاروالو و همکاران [۱۵]، در پژوهشی به تحلیل احساسات در گفتگوها پرداختند، با تمرکز بر تعاملات انسان-ماشین و انسان-انسان. هدف اصلی این تحقیق بررسی تأثیر بافت بر بهبود عملکرد مدل‌های تحلیل احساسات بود. بافت در این مطالعه شامل جملات قبلی و نقش گوینده (مانند مشتری یا نماینده خدمات) بود. این پژوهش به سوالاتی درباره تأثیر تعداد جملات قبلی، نقش گوینده، و انتخاب مدل مناسب برای تحلیل احساسات در گفتگوها پاسخ داد. برای دستیابی به این هدف، از ترکیبی از روش‌های سنتی مانند رگرسیون لجستیک، SVM، جنگل تصادفی و CRF؛ و مدل‌های پیشرفته یادگیری عمیق از جمله BERT، رویکرد BERT بهینه شده قوی[3] (RoBERTa) و BERT-CRF استفاده شد. همچنین، مدل‌های یادگیری کم فرصت[4] (FSL) مانند GPT-3 و OPT مورد بررسی قرار گرفتند. نتایج این پژوهش نشان داد که مدل‌های BERTimbau و RoBERTa بهترین عملکرد را در تحلیل احساسات با در نظر گرفتن بافت داشتند. در گفتگوهای انسان-ماشین، در نظر گرفتن بافت هر دو گوینده (مشتری و نماینده خدمات) بهبود عملکرد مدل را به همراه داشت. در حالی که در گفتگوهای انسان-انسان، در نظر گرفتن جملات مشتری کافی بود. روش‌های FSL نیز نتایج قابل قبولی ارائه دادند، اما به اندازه مدل‌های مبتنی بر BERT مؤثر نبودند. این پژوهش با محدودیت‌هایی مواجه بود. داده‌ها محدود به زبان پرتغالی بودند و نتایج ممکن است برای زبان‌های دیگر متفاوت باشد. مدل‌های FSL به دلیل هزینه‌های محاسباتی بالا و عملکرد ضعیف‌تر نسبت به BERT محدودیت داشتند. همچنین، برخی مدل‌ها مانند RoBERTa در پردازش جملات طولانی به مشکل برخوردند.

کیان لانگ تن و همکاران [۱۶]، در یکی از پژوهش‌های خود به بررسی و توسعه رویکردی جدید برای تحلیل احساسات و شناسایی و طبقه‌بندی احساسات مثبت، منفی و خنثی پرداخته‌اند. در این پژوهش، تمرکز اصلی بر استفاده از مدل‌های ترکیبی عمیق بوده است که با ترکیب ویژگی‌های مدل‌های مبتنی بر Transformer مانند RoBERTa و مدل‌های دنباله‌ای نظیر LSTM، BiLSTM و GRU بهبود دقت و عملکرد تحلیل احساسات را هدف قرار داده‌اند. در این مطالعه، مدل ترکیبی عمیقی معرفی شده است که شامل سه ساختار مجزا یعنی RoBERTa-LSTM، RoBERTa-BiLSTM و RoBERTa-GRU است. برای ترکیب نتایج این مدل‌های ترکیبی، دو روش میانگین‌گیری و رأی اکثریت به کار گرفته شد. همچنین، برای مقابله با مشکل عدم تعادل داده‌ها، از روش GloVe برای افزایش داده‌ها و ایجاد نمونه‌های بیشتر برای کلاس‌های اقلیت استفاده شده است. نتایج نشان داده که مدل پیشنهادی با استفاده از روش رأی اکثریت توانسته است دقتی معادل ۹۴,۹ درصد در مجموعه داده IMDB، ۹۱,۷۷ درصد در مجموعه داده Twitter US Airline Sentiment و ۸۹,۸۱ درصد در مجموعه داده Sentiment140 به دست آورد. در این میان، استفاده از RoBERTa به عنوان لایه ابتدایی، عملکردی برتر نسبت به مدل‌های مشابه مانند BERT و ALBERT نشان داده است. با این وجود، این رویکرد با چالش‌هایی نیز روبروست. نیازمندی بالای منابع محاسباتی به دلیل استفاده از مدل‌های پیچیده مانند RoBERTa و مدل‌های دنباله‌ای، یکی از مهم‌ترین محدودیت‌های این روش است. علاوه بر این، افزایش داده‌ها با استفاده از GloVe گاهی ممکن است منجر به تولید نمونه‌های مصنوعی با کیفیت پایین شود.

کیان لانگ تن و همکاران [۱۷] همچنین پژوهشی قدیمی‌تر، با هدف بهبود دقت تحلیل احساسات ارائه کرده‌اند که در آن مدلی ترکیبی از قابلیت‌های مدل RoBERTa و LSTM بهره می‌گیرد. این مدل تلاش می‌کند تا با حل چالش‌هایی مانند وابستگی‌های بلندمدت در متن و تنوع واژگانی، عملکردی قابل‌توجه در تحلیل احساسات به نمایش بگذارد. در این راستا، از روش‌های پیشرفته‌ای همچون جاسازی‌های معنایی قدرتمند، تکنیک‌های افزایش داده برای مقابله با عدم توازن داده‌ها و پیش‌پردازش دقیق متن استفاده شده است. در این روش، مدل RoBERTa به‌عنوان ابزاری برای تبدیل متن به بازنمایی‌های عددی معنادار عمل می‌کند؛ در حالی که LSTM با قابلیت پردازش وابستگی‌های بلندمدت، ویژگی‌های زمانی متن را استخراج و تحلیل می‌کند. این بازنمایی‌های معنایی و زمانی در یک ساختار یکپارچه ادغام شده و با استفاده از لایه‌های تکمیلی نظیر Flatten و Dense به طبقه‌بندی نهایی منجر می‌شوند.

[1] Support Vector Machine
[2] Extreme Learning Machine
[3] Robustly Optimized BERT Approach
[4] Few-Shot Learning

فرآیند آموزش مدل نیز با بهره‌گیری از بهینه‌ساز برآورد لحظه تطبیقی[1] (Adam) و تابع زیان متقاطع طبقه‌ای [2] بهینه‌سازی شده است. نتایج به‌دست‌آمده نشان می‌دهند که مدل ترکیبی RoBERTa-LSTM در مقایسه با روش‌های سنتی مانند نایو بیز، SVM و LSTM عملکرد بهتری ارائه داده و توانسته است در مجموعه داده‌های مختلف به دقت بالایی دست یابد. تکنیک‌های افزایش داده نیز نقش کلیدی در بهبود دقت مدل به‌ویژه در مواجهه با داده‌های نامتوازن ایفا کرده‌اند و بهبود قابل‌توجهی در معیارهایی همچون امتیاز F1 ایجاد کرده‌اند. با وجود این دستاوردها، مدل پیشنهادی با محدودیت‌هایی نیز روبه‌رو است. از جمله این محدودیت‌ها می‌توان به نیاز به منابع محاسباتی بالا و زمان طولانی‌تر برای آموزش به دلیل پیچیدگی ساختار مدل اشاره کرد.

مهمت و ایلهان [۸]، مهمت و ایلهان در پژوهش خود به طراحی یک مدل ترکیبی نوآورانه برای طبقه‌بندی احساسات پرداخته‌اند که از ترکیب روش‌های مختلف یادگیری عمیق و تکنیک‌های تعبیه کلمات بهره می‌برد. هدف آن‌ها بهبود دقت تحلیل احساسات، به‌ویژه در بررسی توییت‌های ترکی، بوده است. مدل پیشنهادی آن‌ها از روش‌های متنوع تعبیه کلمات، از جمله Word2Vec، FastText و تعبیه در سطح کاراکتر، به همراه مدل‌های یادگیری عمیق نظیر CNN، LSTM، BiLSTM و GRU استفاده کرده است. نوآوری کلیدی در این پژوهش، ترکیب موازی CNN و BiLSTM با بهره‌گیری از تعبیه در سطح کاراکتر و FastText است که خروجی‌های این دو بخش پس از ترکیب، به یک لایه سافت مکس برای طبقه‌بندی نهایی منتقل می‌شوند. در این پژوهش، پیش‌پردازش داده‌ها با دقت بالایی انجام شده است و شامل مراحلی نظیر تبدیل حروف به حروف کوچک، حذف علائم نگارشی و اعداد، تصحیح اشتباهات املایی با ابزار Zemberek، و جایگزینی لینک‌ها و نام‌های کاربری با برچسب‌های خاص بوده است. نتایج به‌دست‌آمده نشان داده است که مدل ترکیبی پیشنهادی، با دقت ۸۲,۱۴٪ عملکرد بهتری نسبت به سایر مدل‌های پایه ارائه کرده است. برای مثال، مدل CNN با تعبیه در سطح کاراکتر دقت ۷۵,۶۷٪ و مدل BiLSTM با FastText دقت ۸۰,۴۴٪ داشته‌اند. با این حال، پژوهش آن‌ها با محدودیت‌هایی نظیر کمبود داده، دشواری در مدیریت اصطلاحات خاص و طعنه، و اهمیت یکسان برای تمام کلمات روبه‌رو بوده است.

سئونگ وان سو و همکاران [۱۸] پژوهشی جامع با هدف مقایسه مدل‌های مبتنی بر یادگیری عمیق در طبقه‌بندی احساسات ارائه داده‌اند. در این مطالعه، عملکرد مدل‌های مختلف CNN و RNN با تأکید بر دو نوع ورودی سطح کلمه و سطح کاراکتر مورد بررسی قرار گرفته است. این پژوهش شامل آزمایش مدل‌ها بر روی ۱۳ مجموعه‌داده متنوع بوده و تلاش کرده است نقش ویژگی‌های مجموعه‌داده، ساختار مدل، و نوع ورودی را در کارایی طبقه‌بندی احساسات روشن کند. پژوهشگران پرسش‌هایی کلیدی را مطرح کرده‌اند، از جمله اینکه تأثیر ویژگی‌های مجموعه‌داده بر تحلیل احساسات چگونه است؟ ساختارهای CNN و RNN چه نقشی در بهبود طبقه‌بندی احساسات دارند؟ و چگونه نوع ورودی می‌تواند عملکرد مدل‌ها را ارتقا بخشد؟ به‌منظور پاسخ به این پرسش‌ها، هشت مدل مختلف آزمایش شده‌اند که شامل CNN ساده با یک لایه کانولوشنال، CNN نه‌لایه، و CNN بسیار عمیق با ۲۹ لایه هستند. همچنین مدل‌های RNN شامل RNN پایه‌ای، LSTM، GRU، و نسخه‌های دوطرفه LSTM و GRU که قادر به پردازش اطلاعات از هر دو جهت هستند، مورد بررسی قرار گرفته‌اند. تمامی این مدل‌ها با ورودی‌های سطح کلمه و کاراکتر ارزیابی شده‌اند. نتایج این پژوهش نشان داده است که افزایش حجم داده‌های آموزشی به بهبود قابل‌توجه عملکرد مدل‌ها منجر می‌شود. مدل‌های RNN، به‌ویژه با ورودی سطح کلمه، در اکثر موارد عملکرد بهتری نسبت به CNN نشان داده‌اند. در میان این مدل‌ها، نسخه‌های دوطرفه LSTM و GRU بهترین نتایج را ارائه داده‌اند. با این حال، CNN با ورودی سطح کاراکتر در برخی موارد، به‌ویژه در ساختارهای عمیق‌تر، عملکرد بهتری داشته است. این پژوهش دارای نقاط قوت متعددی است، از جمله بررسی جامع مدل‌های CNN و RNN، استفاده از ۱۳ مجموعه‌داده متنوع برای تضمین قابلیت تعمیم‌پذیری، و ارائه بینش‌هایی عملی برای انتخاب مدل‌های مناسب تحلیل احساسات. با این وجود، محدودیت‌هایی نیز وجود دارد. برای مثال، این مطالعه تنها از بردارهای کلمه و کاراکتری که از ابتدا آموزش داده شده‌اند استفاده کرده و بردارهای از پیش آموزش‌دیده مانند BERT را در نظر نگرفته است. همچنین، ترکیب مدل‌های CNN و RNN و تأثیر ویژگی‌های زبانی خاص نظیر ساختارهای نحوی و معنایی بر عملکرد مدل‌ها بررسی نشده است.

نینگ ژینگ و همکاران [۱۹]، در پژوهشی جامع به توسعه و ارزیابی یک مدل یادگیری چندوظیفه‌ای برای تحلیل احساسات در نظرات کاربران درباره کالاهای مختلف پرداختند. این پژوهش با هدف ایجاد مدلی که بتواند ویژگی‌های محلی و جهانی متن را به طور همزمان

[1] Adaptive Moment Estimation

[2] Categorical Crossentropy

و مؤثر ترکیب کند، طراحی شده است. در مدل پیشنهادی، از ترکیب CNN چندمقیاسی برای استخراج ویژگی‌های محلی و LSTM برای شناسایی ویژگی‌های دنباله‌ای و جهانی استفاده شده است. این دو دسته ویژگی در نهایت در لایه‌ای موسوم به Fusion-Net ادغام شده‌اند تا نمایش یکپارچه و جامعی از متن به دست آید. یکی از ویژگی‌های کلیدی این مدل، استفاده از یادگیری چندوظیفه‌ای است که در آن، یک رمزگذار مشترک وظیفه استخراج ویژگی‌های کلی را بر عهده دارد؛ در حالی که رمزگذارهای خصوصی برای استخراج ویژگی‌های خاص هر نوع کالا به کار می‌روند. علاوه بر این، مکانیزم یادگیری خصمانه[1] در ساختار مدل تعبیه شده است تا از استقلال ویژگی‌های مشترک از نوع کالا اطمینان حاصل شود و تداخل اطلاعات میان ویژگی‌های مشترک و خصوصی به حداقل برسد. نتایج این پژوهش نشان داد که مدل MTL-MSCNN-LSTM با دستیابی به دقت ۸۷٫۳۳ درصد و امتیاز F1 معادل ۰٫۸۷۳۴ عملکرد قابل‌توجهی در مقایسه با روش‌های سنتی نظیر رگرسیون لجستیک، SVM و جنگل تصادفی داشته است. این یافته‌ها همچنین نشان داد که یادگیری چندوظیفه‌ای به‌طور معناداری عملکرد مدل را نسبت به یادگیری تک‌وظیفه‌ای بهبود داده است. ترکیب ویژگی‌های محلی و جهانی از طریق Fusion-Net نیز تأثیر بسزایی در افزایش دقت مدل داشته است. با این حال، این مدل با محدودیت‌هایی نیز روبه‌روست. از جمله این که ساختار پیچیده آن نیازمند منابع محاسباتی بیشتری نسبت به روش‌های سنتی است و زمان آموزش مدل نیز به طور قابل‌توجهی افزایش می‌یابد. علاوه بر این، آزمایش‌ها تنها بر داده‌های مربوط به نظرات کاربران درباره کالاها انجام شده و کارایی مدل در سایر حوزه‌ها بررسی نشده است.

رونگ زنگ و همکاران [۲۰] در پژوهشی تلاش کردند تا چالش‌های مربوط به طبقه‌بندی احساسات بین دامنه‌ای[2] را برطرف کنند؛ مشکلی که ناشی از تفاوت‌های ساختاری و معنایی میان دامنه‌های مختلف است. به این منظور، آن‌ها روشی ترکیبی مبتنی بر CNN و یادگیری گسترده طراحی کردند که قادر است به‌طور هم‌زمان ویژگی‌های مشترک میان دامنه‌ها و ویژگی‌های خاص هر دامنه را استخراج کند. این رویکرد باعث شد که عملکرد طبقه‌بندی احساسات در دامنه‌های مختلف بهبود یابد. علاوه بر این، از ابزارهایی نظیر حداکثر اختلاف میانگین[3] (MMD) برای کاهش اختلاف توزیع داده‌ها و Co-Training برای بهره‌برداری بهینه از داده‌های محدود برچسب‌گذاری‌شده در دامنه هدف استفاده شده است. مدل پیشنهادی آن‌ها توانست از طریق ترکیب CNN و یادگیری گسترده، ویژگی‌های محلی و الگوهای معنایی عمیقی از متن استخراج کند. بهره‌گیری از BERT به‌عنوان ابزار بازنمایی معنایی عمیق، قدرت مدل را در درک متن‌های پیچیده افزایش داد. در همین راستا، MMD نیز با کاهش شکاف دامنه‌ای و انتقال دانش میان دامنه‌ها، توانایی مدل را در انطباق با داده‌های دامنه هدف بهبود بخشید. در نهایت، Co-Training امکان بهره‌برداری بهتر از داده‌های محدود دامنه هدف را فراهم آورد و به بهبود دقت مدل کمک کرد. با این وجود، این پژوهش نیز با محدودیت‌هایی همراه بوده است. نیاز به منابع محاسباتی بالا و کمبود داده‌های برچسب‌گذاری‌شده در دامنه هدف از چالش‌های اصلی مدل پیشنهادی هستند. علاوه بر این، پیچیدگی‌های زبانی و محدودیت در پردازش متن‌های طولانی و پیچیده، گاهی منجر به کاهش عملکرد مدل شده است.

محمدرضا فیضی درخشی و همکاران [۲۱] پژوهشی با هدف توسعه مدلی نوآورانه برای تحلیل احساسات چندوجهی انجام داده‌اند که تلاش دارد با بهره‌گیری هم‌زمان از ویژگی‌های متنی و تصویری، دقت در طبقه‌بندی احساسات را بهبود بخشد. روش پیشنهادی آنان بر پایه شبکه‌های عمیق توجه‌محور طراحی شده است تا تعامل مؤثر میان این دو نوع ویژگی را تقویت کرده و بازنمایی معنایی عمیق‌تری از داده‌ها ارائه دهد. علاوه بر این، از روش‌های تفسیرپذیر مانند مدل تفسیری محلی- مدل توضیحی آگنوستیک[4] (LIME) استفاده شده تا تصمیمات مدل برای کاربران شفاف و قابل‌درک باشد، امری که به اعتماد بیشتر به مدل کمک می‌کند. این مدل از شبکه‌های عصبی توجه چندوجهی عمیق[5] (DMVAN) برای استخراج ویژگی‌های تصویری و متنی از چندین سطح و دیدگاه بهره می‌گیرد. مکانیزم‌های توجه چندوجهی در این مدل، تعامل میان این ویژگی‌ها را بهبود می‌بخشند، به گونه‌ای که بتوان ویژگی‌های احساسی تصاویر را با کمک متن بهتر شناسایی کرد. برای ادغام این ویژگی‌ها در یک ساختار جامع، از رویکرد ادغام چندوجهی همراه با مکانیزم‌های توجه چندسر استفاده شده است که تحلیل و بازنمایی دقیق‌تری از اطلاعات ارائه می‌دهد. همچنین، لایه‌های شبکه عصبی چندلایه[6] (MLP) به مدل افزوده شده تا عمق تحلیل افزایش یابد و دقت در طبقه‌بندی بهبود پیدا کند. نتایج این پژوهش نشان می‌دهد که مدل پیشنهادی عملکردی بهتر از روش‌های پیشین داشته است. این موفقیت به دلیل استخراج بازنمایی‌های معنایی عمیق‌تر و تعامل مؤثرتر میان ویژگی‌های متنی و تصویری است. افزون بر این،

[1] Adversarial Learning
[2] Cross-Domain Emotion Classification
[3] Maximum Mean Discrepancy
[4] local interpretable model-agnostic explanation model
[5] deep multi-view attentive network
[6] multi-layer perceptron

بهره‌گیری از روش تفسیرپذیر LIME به تحلیل شفاف‌تر تصمیمات مدل کمک کرده است، که این امر باعث افزایش اعتماد به کاربرد آن می‌شود. با این حال، نیاز به منابع محاسباتی قوی و کمبود داده‌های برچسب‌گذاری‌شده برای تحلیل احساسات چندوجهی از جمله چالش‌های اصلی پژوهش هستند. همچنین، پیچیدگی‌های زبانی و دشواری در پردازش متن‌های طولانی‌تر و پیچیده‌تر، از دیگر محدودیت‌های موجود به شمار می‌روند.

فالیان هوانگ و همکاران [۲] مدلی نوین با نام AEC-LSTM معرفی کرده‌اند که به‌منظور بهبود تحلیل احساسات متنی طراحی شده است. این مدل با ترکیب مکانیزم‌های توجه، هوش هیجانی، و شبکه‌های عصبی LSTM و CNN توانسته است گامی مؤثر در استخراج و تحلیل ویژگی‌های احساسی متن بردارد. نقطه تمرکز اصلی این مدل، بهره‌گیری از هوش هیجانی برای تقویت فرآیند یادگیری ویژگی‌های احساسی و استفاده از مکانیزم توجه مبتنی بر موضوع برای تنظیم وزن‌های بازنمایی پنهان متن است. این مکانیزم، ویژگی‌های احساسی مرتبط با موضوع را به شکلی دقیق‌تر و هدفمندتر استخراج می‌کند. در مدل AEC-LSTM، یک نسخه پیشرفته از LSTM به‌کار گرفته شده که هوش هیجانی را به طور مستقیم در فرآیند یادگیری اطلاعات احساسی دخیل می‌کند. این کار با استفاده از یک ماژول مبتنی بر هیجان و تخمین‌زننده هیجان انجام می‌شود که یادگیری ویژگی‌های احساسی را بهبود می‌بخشد. در همین راستا، مکانیزم توجه مبتنی بر موضوع به مدل امکان می‌دهد تا با تمرکز بر موضوعات خاص متن، بازنمایی‌های دقیق‌تر و غنی‌تری از احساسات ارائه دهد. علاوه بر این، CNN برای استخراج ویژگی‌های محلی متن طراحی شده‌اند و ترکیب آن‌ها با LSTM، امکان تحلیل وابستگی‌های بلندمدت و ویژگی‌های محلی را به‌صورت هم‌زمان فراهم کرده است. نتایج آزمایش‌های انجام‌شده بر روی مجموعه داده‌های واقعی مانند IMDB، Yelp2014، JDReview و SinaWeibo نشان می‌دهد که مدل AEC-LSTM در مقایسه با روش‌های پیشین عملکرد بهتری دارد. این مدل به لطف استفاده از مکانیزم‌های توجه و هوش هیجانی، دقت طبقه‌بندی احساسات را به طور قابل‌توجهی افزایش داده است. همچنین، ترکیب توانمندی‌های LSTM و CNN، این مدل را به ابزاری قدرتمند برای تحلیل احساسات تبدیل کرده است. با این حال، این مدل با محدودیت‌هایی نیز مواجه است. نیاز به منابع محاسباتی بالا و وابستگی به داده‌های برچسب‌گذاری‌شده از جمله چالش‌های اصلی آن محسوب می‌شود. همچنین، پیچیدگی ساختاری مدل می‌تواند منجر به افزایش زمان آموزش و پیش‌بینی شود.

مین دونگ و همکاران [۴] در این پژوهش مدلی جدید با نام شبکه عصبی کانولوشن متغیر و ادغام کانولوشن[1] (VCPCNN) معرفی کرده‌اند که به‌منظور بهبود دقت طبقه‌بندی احساسات متنی طراحی شده است. هدف این مدل، بهینه‌سازی فرآیندهای کانولوشن و ادغام با استفاده از معماری‌های پیشرفته‌ای است که بر محدودیت‌های موجود در TextCNN غلبه می‌کنند. در این راستا، مدل VCPCNN از لایه‌های کانولوشنی متغیر و ترکیبی از روش‌های ادغام میانگین و حداکثر بهره می‌برد تا ویژگی‌های احساسی متن را با دقت بیشتری استخراج کند. این تغییرات ساختاری، امکان حفظ اطلاعات جزئی و حیاتی در بازنمایی‌های استخراج‌شده را فراهم می‌کنند و به بهبود قابل‌توجه عملکرد مدل منجر شده‌اند. در این پژوهش، عملکرد VCPCNN با مدل‌های دیگری مانند TextCNN، MVCNN، RCNN و BiLSTM مقایسه شده است. نتایج به‌دست‌آمده حاکی از بهبود عملکردی معادل ۱,۹۷ درصد نسبت به TextCNN در چهار مجموعه داده متفاوت است. همچنین، در تحلیل داده‌های چنددسته‌ای چینی، VCPCNN توانسته است دقت بیشتری نسبت به مدل‌های پیشین ارائه دهد و سازگاری بهتری با داده‌های پیچیده و دسته‌بندی‌های متنوع نشان دهد. این موفقیت، توانایی بالای مدل در شناسایی ویژگی‌های حساس و مهم متن را به‌وضوح نشان می‌دهد. با این حال، مقاله به محدودیت‌هایی نیز اشاره دارد. یکی از این محدودیت‌ها، ضعف TextCNN در استخراج مناسب ویژگی‌های تعبیه کلمات است، که این مشکل در بسیاری از مدل‌های مشابه نیز دیده می‌شود. نویسندگان بر این باورند که مدل‌های موجود اغلب اطلاعات مربوط به ابعاد تعبیه کلمات را نادیده می‌گیرند و نیاز به بهبود در این زمینه کاملاً محسوس است.

ژو لیانگ لنگ و همکاران [۲۲] در پژوهش خود مدلی ترکیبی برای بهبود تحلیل احساسات متنی ارائه کرده‌اند که از ترکیب شبکه‌های عصبی بازگشتی نظیر BiLSTM و BiGRU با مکانیزم خودتوجهی چند سر تقویت شده[2] بهره می‌برد. هدف این مدل، تقویت بازنمایی معنایی متن‌ها و بهبود دقت و عملکرد در تحلیل احساسات است. همچنین، استفاده از ساختار رمزگذار خودکار به بازتولید اطلاعات متنی و ارتقای دقت طبقه‌بندی کمک می‌کند. مکانیزم خودتوجهی چند سر تقویت شده نقش کلیدی در این مدل دارد؛ که از ساختار Transformer الهام گرفته شده است. این مکانیزم با حذف عملیات ماسک‌کردن و لایه‌های پیش‌خور، بر تعاملات معنایی بین کلمات تمرکز بیشتری دارد و بازنمایی‌های دقیق‌تر و معنایی‌تری ارائه

[1] variable convolution and pooling convolution neural network

[2] Enhanced Multi-Head Self-Attention

می‌دهد. در همین راستا، BiLSTM و BiGRU به‌عنوان لایه‌های رمزگذار و رمزگشا، به استخراج وابستگی‌های زمانی و بازنمایی ویژگی‌های عمیق کمک می‌کنند. همچنین، استفاده از BERT در پیش‌پردازش به جای Word2Vec، موجب تولید جاسازی‌هایی با دقت معنایی بالاتر شده است. نتایج این پژوهش نشان داده است که مدل RNN هیبریدی و خودتوجهی چند سر پیشرفته[1] (HRNaEMSA) نسبت به روش‌های پایه نظیر Text-CNN، LSTM-Attention و RCNN در معیارهایی مانند دقت، صحت[2]، یادآوری[3] و امتیاز F1 عملکرد بهتری دارد. به‌طور خاص، مکانیزم خودتوجهی چند سر تقویت شده به بازنمایی دقیق‌تر ویژگی‌های احساسی متن و افزایش دقت طبقه‌بندی کمک کرده است. بررسی‌ها همچنین نشان داده که BiLSTM در مقایسه با BiGRU از نظر دقت برتر است، اما BiGRU از لحاظ سرعت و کاهش تعداد پارامترها عملکرد بهتری دارد. علاوه بر این، بهره‌گیری از BERT در پیش‌پردازش، برتری قابل‌توجهی نسبت به مدل‌های پایه مبتنی بر BERT ایجاد کرده است. با وجود این دستاوردها، مدل HRNaEMSA با محدودیت‌هایی نیز روبه‌رو است. این مدل تنها بر روی دو مجموعه داده IMDB و SST-2 آزمایش شده و نیاز به ارزیابی در مجموعه داده‌های بزرگ‌تر و متنوع‌تر دارد. علاوه بر این، پیچیدگی مدل و تعداد بالای پارامترها منجر به افزایش زمان محاسباتی می‌شود و مکانیزم‌های ماسک‌کردن همچنان نیازمند بهینه‌سازی بیشتری هستند.

محمد رضوان رشید رعنا و همکاران [۲۳] در پژوهش خود روشی ترکیبی و نوآورانه برای ABSA ارائه کرده‌اند که از مدل‌های پیشرفته یادگیری عمیق بهره می‌برد. این روش از ترکیب RoBERTa، CNN تک بعدی و BiLSTM برای استخراج ویژگی‌ها و طبقه‌بندی احساسات در نظرات کاربران استفاده می‌کند. هدف اصلی این رویکرد، بهبود دقت و کارایی تحلیل احساسات در نظرات کوتاه و غیرساختاریافته است. در این پژوهش، مدل RoBERTa برای استخراج ویژگی‌های متنی پیشرفته و ایجاد بردارهای تعبیه به کار گرفته شده است. سپس، CNN تک بعدی برای شناسایی وابستگی‌های محلی در متن و BiLSTM برای درک وابستگی‌های متوالی و بلندمدت به‌منظور طبقه‌بندی احساسات استفاده شده است. روش ترکیبی پیشنهادی، قدرت استخراج ویژگی‌های پیشرفته توسط RoBERTa را با توانایی CNN تک بعدی در شناسایی الگوهای محلی و قابلیت BiLSTM در تحلیل روابط بلندمدت ترکیب می‌کند. نتایج این پژوهش نشان داده است که مدل پیشنهادی با دقت ۹۲,۳۳ درصد در سه مجموعه داده استاندارد شامل Hu and Liu، SemEval و Bo Pang and Lillian Lee ارزیابی شده و عملکرد بهتری نسبت به روش‌های موجود مانند BERT-ADA، MGAN و LMIAN داشته است. این مدل به‌ویژه در تحلیل نظرات کوتاه و غیرساختاریافته عملکرد موفقی از خود نشان داده است. استفاده از RoBERTa برای درک معنایی دقیق‌تر متن و ترکیب آن با CNN تک بعدی و BiLSTM، دقت و تفسیرپذیری نتایج را به‌طور قابل توجهی بهبود بخشیده است. با این حال، مدل پیشنهادی با محدودیت‌هایی نیز مواجه است. برای مثال، در تحلیل داده‌های بسیار پیچیده یا نظرات دارای جنبه‌های متعدد، دقت مدل ممکن است کاهش یابد. همچنین، آموزش مدل نیازمند منابع محاسباتی بالایی است که ممکن است در شرایط محدودیت منابع به چالش تبدیل شود.

بوآسیدا یوسرا و مزالی حکیم [۲۴] در پژوهش خود به بررسی و توسعه مدل‌های ترکیبی برای تحلیل احساسات در پلتفرم‌های اجتماعی نظیر توییتر پرداخته‌اند. هدف اصلی این تحقیق، بهبود تحلیل احساسات از طریق ترکیب مدل‌های Transformer و مدل‌های دنباله‌ای مانند BiLSTM بوده است. این پژوهش تلاش کرده تا شکاف میان توانایی‌های مدل‌های Transformer در درک مفهومی متن و نقاط قوت مدل‌های دنباله‌ای در پردازش وابستگی‌های زمانی را پر کند. در این مطالعه، سه مدل ترکیبی جدید معرفی شده است: RoBERTa-CNN-BiLSTM که از RoBERTa برای استخراج مفاهیم، CNN برای استخراج ویژگی‌های محلی و BiLSTM برای پردازش وابستگی‌های دنباله‌ای بهره می‌برد؛ BERT-BiLSTM که ترکیب BERT برای ایجاد تعبیه‌های مفهومی و BiLSTM برای تحلیل دنباله‌ای است؛ و DistilBERT-BiLSTM که با استفاده از نسخه سبک‌تر BERT و BiLSTM به کاهش منابع محاسباتی و افزایش دقت پرداخته است. داده‌های این پژوهش با مجموعه‌ی Sentiment140، شامل ۱,۶ میلیون توییت با برچسب‌های مثبت و منفی، ارزیابی شده‌اند و پیش‌پردازش‌هایی مانند توکن‌سازی، حذف کاراکترهای خاص و تبدیل متن به تعبیه‌های مدل Transformer نیز انجام گرفته است. نتایج نشان داده‌اند که مدل DistilBERT-BiLSTM با دقت ۸۱ درصد و معیارهای صحت و یادآوری به ترتیب ۸۲ و ۸۰ درصد، بهترین عملکرد را داشته است. مدل‌های BERT-BiLSTM و RoBERTa-CNN-BiLSTM نیز به ترتیب دقت‌های ۷۹ و ۷۷ درصد را ارائه داده‌اند. این یافته‌ها نشان‌دهنده بهبود قابل توجه مدل‌های ترکیبی

[1] Hybrid RNN and Enhanced Multi- Head Self-Attention
[2] Precision
[3] Recall

نسبت به روش‌های سنتی در تحلیل احساسات توییتر هستند. از نقاط قوت این تحقیق می‌توان به ترکیب قابلیت‌های Transformer و BiLSTM، استفاده از DistilBERT به‌عنوان مدلی سبک‌تر برای کاهش نیاز به منابع محاسباتی و بهبود دقت تحلیل احساسات در داده‌های پیچیده و غیرساختاریافته اشاره کرد. با این حال، این پژوهش محدودیت‌هایی نیز دارد؛ برای مثال، مدل‌ها تنها بر روی مجموعه داده Sentiment140 ارزیابی شده‌اند و عملکرد آنها بر روی داده‌های دیگر نیازمند بررسی است. همچنین، منابع محاسباتی نسبتاً بالایی برای آموزش مدل‌ها نیاز است. این پژوهش محدودیت‌هایی مانند نیاز به تحلیل داده‌های چندزبانه و پیچیده‌تر، توسعه مدل‌های ترکیبی برای تحلیل احساسات در زمان واقعی، و ارتقای مدل‌ها برای داده‌های با جنبه‌های متعدد و پیچیده را شناسایی کرده است.

پژوهش ما با در نظر گرفتن تحقیقات پیشین و تمرکز بر تحلیل احساسات در زبان فارسی، با هدف رفع محدودیت‌های موجود، کاربردهای گسترده‌ای را برای تحلیل احساسات در شرایط کم‌منبع و زبان‌های دیگر، از جمله زبان عربی با رسم‌الخط مشابه، فراهم می‌کند. جدول۱، خلاصه‌ای از توضیحات عملکرد مدل‌های پیشنهادی در پیشینه را ارائه می‌دهد.

جدول ۱ خلاصه ارزیابی پژوهش‌های پیشین

افراد مرتبط به مدل	نام مدل	توضیحات، مزایا و معایب مدل	عملکرد مدل
فضلی واحید و همکاران [۱۲]	SAE- LSTM	مزایا: دقت مدل ترکیبی با نسبت ۹۰ درصد آموزش و ۱۰ درصد آزمون، ۸۷ درصد را ثبت کرد که بهتر از مدل‌های یادگیری عمیق ساده است. معایب: عدم بررسی طبقه‌بندی چند طبقه‌ای که بی‌طرفی یا دوسوگرایی را نادیده می‌گیرد.	دقت ٪۸۷
مریگانک شوکلا و آخیل کومار [۶]	Transformer-BiLSTM-CNN	مزایا: دقت بالاتر و عملکرد بهتر نسبت به مدل‌های ساده مانند LSTM، BiLSTM و CNN، استفاده از لایه رمزگذار Transformer، دقت را تا ۲ الی ۳ درصد افزایش می‌دهد. معایب: پیچیدگی بالای مدل، ترکیب CNN و BiLSTM، دقت مدل CNN را افزایش چندانی نمی‌دهد.	دقت ٪۸۹,۰۴
سکرین تام و همکاران [۱]	ConvBiLSTM	بهبود عملکرد طبقه‌بندی احساسات نسبت به مدل‌های ساده‌ی CNN، LSTM، Bi-LSTM و CNN-LSTM با تعبیه کلمه‌ی Word2Vec معایب: داده‌های مورد استفاده محدود بوده‌اند که قابل تعمیم به دنیای واقعی نیستند.	دقت ٪۹۴,۱۳
شفیق و همکاران [۱۳]	AraBERT-CRF	مزایا: بعد از تنظیم پارامترهای مدل AraBERT برای E2E-ABSA، AraBERT با CRF در طبقه‌بندی بهتر از سافت مکس عمل می‌کند. معایب: نیازمند منابع محاسباتی قدرتمند از جمله پردازنده‌های قوی GPU، نیازمند داده آموزشی بیشتر و متنوع‌تری برای رسیدن به عملکرد بهینه. استفاده از CRF فضای جستجو را بزرگتر و پیچیده‌تر می‌کند.	دقت ٪۹۸,۳۴ امتیاز F1 ٪۹۷,۷۸
محمد الجبرین و همکاران [۱۴]	MFOHDL-SA	مزایا: مدل پیشنهادی نسبت به رویکردهای SVM، XGBoost و CNN بهبود یافته است. ترکیب الگوریتم MFO و CNN-LSTM باعث افزایش دقت طبقه‌بندی شد. معایب: پیچیده بودن الگوریتم MFO و پیاده‌سازی چالش‌برانگیز	دقت ٪۹۵,۰۹ امتیاز F1 ٪۹۲,۶۲
ایزابل کاروالو و همکاران [۱۵]	BERTimbau	مزایا: دقت بالاتر در طبقه‌بندی احساسات نسبت به مدل‌های SVM، نایو بیز و RNN ساده، در نظر گرفتن بافت بیشتر با استفاده از CRF معایب: پیچیدگی محاسباتی و زمان آموزش بالاتر نسبت به مدل‌های ساده‌تر، نیاز به تنظیم تعداد پارامترهای بیشتری نسبت به SVM و نایو بیز، افت عملکرد با افزایش بیش از حد طول جمله‌ها.	امتیاز F1 ٪۸۴
کیان لانگ تن و همکاران [۱۶]	ensemble hybrid RoBERTa-LSTM (Majority)	مزایا: ترکیب چندین مدل یادگیری عمیق و استفاده از تکنیک‌های مجموعه موجب بهبود عملکرد، نسبت به RoBERTa-LSTM شد. معایب: پیچیدگی بیشتر نسبت به یک مدل تکی	دقت ٪۹۴,۹
کیان لانگ تن و همکاران [۱۷]	RoBERTa-LSTM	مزایا: استفاده از LSTM موجب یادگیری معنای دنباله دار در جملات طولانی می شود. معایب: تنها LSTM مورد استفاده قرار گرفت، سایر مدل‌های توالی نظیر BiLSTM و GRU مورد بررسی قرار نگرفته‌اند و این مدل تکی، عملکرد ضعیف تری نسبت به مدل مجموعه ای ثبت کرد.	دقت ٪۹۲,۹۶ امتیاز F1 ٪۹۳
مهمت و ایلهان [۸]	M-Hybrid CNN، (ترکیب LSTM، و BiLSTM و GRU)	مزایا: ترکیب تعبیه‌ها با مدل‌های یادگیری عمیق عملکرد طبقه‌بندی احساسات را بهبود می‌بخشد، همچنین دقت مدل پیشنهادی از مدل‌های پایه CNN، LSTM، BiLSTM بالاتر بود. معایب: مجموعه داده‌های مورد استفاده در این مطالعه خیلی بزرگ نیست، صفت ها و قیدها از نظر احساسی دارای اهمیت بیشتری نسبت به اسم‌ها دارند.	دقت ٪۸۲,۱۴

افراد مرتبط به مدل	نام مدل	توضیحات، مزایا و معایب مدل	عملکرد مدل
سئونگ وان سو و همکاران [۱۸]	Bi-LSTM with word-level inputs	مزایا: بررسی جامع مدل‌های مختلف CNN و RNN با دو نوع ورودی. استفاده از ۱۳ مجموعه داده مختلف برای اطمینان از نتایج قابل تعمیم. معایب: استفاده از بردارهای کلمه و کاراکتر که از ابتدا آموزش داده شده‌اند و عدم استفاده از بردارهای از پیش آموزش دیده مانند BERT. عدم بررسی تأثیر ترکیب مدل‌های CNN و RNN در یک معماری ترکیبی.	AUROC (۰,۹۲ - ۰,۶۹)
نینگ ژینگ و همکاران [۱۹]	MTL-MSCNN-LSTM	مزایا: بهبود دقت طبقه‌بندی احساسات با استفاده از استخراج ویژگی‌های محلی و جهانی به طور همزمان. معایب: مدل پیشنهادی نیاز به منابع محاسباتی بیشتری نسبت به روش‌های سنتی مانند SVM و نایو بیز دارد. زمان آموزش مدل به دلیل پیچیدگی ساختار شبکه‌های عمیق بیشتر است.	دقت ٪۸۷,۳۳ امتیاز F1 ٪۸۷,۳۴
رونگ زنگ و همکاران [۲۰]	CNN-CBL) Broad learnig(	مزایا: استفاده از MMD برای کاهش شکاف دامنه‌ای و بهبود انتقال دانش بین دامنه‌ها. بهبود عملکرد طبقه‌بندی احساسات در دامنه‌های مختلف با استفاده از Co-Training. معایب: نیاز به منابع محاسباتی قوی برای آموزش مدل‌های ترکیبی. کمبود داده‌های برچسب‌گذاری شده در دامنه هدف. پیچیدگی‌های زبانی خاص که ممکن است بر عملکرد مدل تأثیر بگذارد.	دقت ٪۸۶,۴۱
محمدرضا فیضی درخشی و همکاران [۲۱]	DMVAN	مزایا: ترکیب هوشمندانه ویژگی‌های متنی و تصویری برای بهبود طبقه‌بندی احساسات. استفاده از مکانیزم‌های توجه چندوجهی برای استخراج ویژگی‌های احساسی و تشخیصی. معایب: نیاز به منابع محاسباتی قوی برای آموزش مدل‌های ترکیبی. کمبود داده‌های برچسب‌گذاری شده برای تحلیل احساسات چندوجهی. پیچیدگی‌های زبانی خاص که ممکن است بر عملکرد مدل تأثیر بگذارد.	دقت ٪۹۹,۸۰۱
فالیان هوانگ و همکاران [۲]	AEC-LSTM	مزایا: استفاده از هوش هیجانی برای بهبود یادگیری ویژگی‌های احساسی. مکانیزم توجه مبتنی بر موضوع که به مدل کمک می‌کند تا ویژگی‌های احساسی مرتبط با موضوع را بهتر استخراج کند. ترکیب LSTM و CNN که باعث می‌شود مدل هم ویژگی‌های محلی و هم وابستگی‌های بلندمدت را استخراج کند. معایب: نیاز به منابع محاسباتی بالا برای آموزش مدل‌های ترکیبی. وابستگی به داده‌های برچسب‌گذاری‌شده برای آموزش مدل. پیچیدگی مدل که ممکن است باعث افزایش زمان آموزش و پیش‌بینی شود.	دقت ٪۹۲,۹۶ امتیاز F1 ٪۹۳
مین دونگ و همکاران [۴]	VCPCNN	مزایا: مدل VCPCNN بهبود ۱,۹۷٪ نسبت به مدل TextCNN را در چهار مجموعه داده مختلف نشان داد. VCPCNN در مجموعه داده‌های چنددسته‌ای چینی عملکرد بهتری نسبت به مدل‌های پیشین داشته است. معایب: محدودیت‌های مدل TextCNN در استخراج ویژگی‌های تعبیه کلمات به خوبی پوشش داده نشده است. این مقاله نشان می‌دهد که مدل‌های موجود اغلب ویژگی‌های تعبیه کلمات را نادیده می‌گیرند.	دقت ٪۸۸
ژو لیانگ لنگ و همکاران [۲۲]	HRNaEMSA (BiLSTM)	مزایا: استفاده از ساختار autoencoder برای بازتولید اطلاعات متنی. بهبود دقت و عملکرد مدل در تحلیل احساسات. معایب: مدل هنوز نیاز به بهبود در استفاده از مکانیزم‌های ماسک‌کردن دارد. زمان محاسباتی و تعداد پارامترها می‌تواند برای مدل‌های بزرگ‌تر به چالش تبدیل شود. نیاز به آزمایش مدل بر روی مجموعه‌های داده بزرگ‌تر و متنوع‌تر.	دقت ٪۸۷,۴ صحت ٪۸۷,۵ یادآوری ٪۸۷,۴ امتیاز F1 ٪۸۷,۴
محمد رضوان رشید رعنا و همکاران [۲۳]	RoBERTa-1D-CNN-BiLSTM	مزایا: استفاده از RoBERTa برای استخراج ویژگی‌های پیشرفته و بهبود درک معنایی متن. ترکیب CNN یک بعدی و BiLSTM برای استخراج ویژگی‌های محلی و بلندمدت و بهبود طبقه‌بندی احساسات. قابلیت مدل در تحلیل نظرات کوتاه و غیرساختاریافته. معایب: مدل پیشنهادی ممکن است در مواجهه با داده‌های بسیار پیچیده یا نظرات با جنبه‌های متعدد دچار کاهش دقت شود. نیاز به منابع محاسباتی بالا برای آموزش مدل‌های ترکیبی.	دقت ٪۹۲,۳۳ امتیاز F1 ٪۹۱,۴۱
بوآسیدا یوسرا و مزالی حکیم [۲۴]	DistilBERT-BiLSTM	مزایا: ترکیب قابلیت‌های Transformer و مدل‌های دنباله‌ای برای بهبود تحلیل احساسات. استفاده از DistilBERT به عنوان یک مدل سبک‌تر که منابع محاسباتی کمتری نیاز دارد. بهبود دقت در تحلیل احساسات در داده‌های پیچیده و غیرساختاریافته توییتر. معایب: مدل‌ها تنها بر روی یک مجموعه داده (Sentiment140) ارزیابی شدند و ممکن است در داده‌های دیگر عملکرد متفاوتی داشته باشند. نیاز به منابع محاسباتی نسبتاً بالا برای آموزش مدل‌های ترکیبی.	دقت ٪۸۱ صحت ٪۸۲ یادآوری ٪۸۰ امتیاز F1 ٪۸۱

۴- روش تحقیق

در این بخش کارهای انجام شده در این پژوهش به طور مفصل شرح داده شده است.

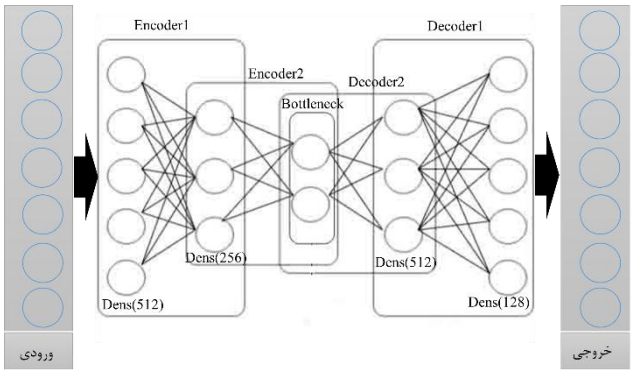
۴-۱- پیش‌پردازش داده‌ها با استفاده از ParsBERT

پیش‌پردازش داده‌ها مرحله‌ای حیاتی در فرآیند یادگیری ماشین است که تأثیر مستقیمی بر عملکرد مدل نهایی دارد. این فرایند شامل مجموعه‌ای از عملیات است که داده‌ها را از حالت خام به قالبی مناسب برای مدل‌سازی تبدیل می‌کند. این فرآیند به‌طور جامع انجام می‌شوند تا داده‌ها برای آموزش مدل‌های یادگیری ماشین به بهترین نحو آماده شوند و دقت و عملکرد مدل نهایی به حداکثر برسد. ابتدا داده‌ها برای اطمینان از کیفیت و صحت آن‌ها پاکسازی

می‌شوند. در این مرحله داده‌های نامعتبر، ناموجود و تکراری حذف می‌شوند. سپس با استفاده از توکنایزر هضم طول کلمات محاسبه می‌شود تا نظرات بسیار کوتاه و بلند شناسایی و مدیریت شود. سپس برای جلوگیری از عدم توازن در داده‌ها، تعداد نظرات مثبت و منفی برابر می‌شوند. در ادامه داده‌ها به سه بخش آموزش، عتبارسنجی و آزمون تقسیم می‌شوند که به ترتیب ۸۱٪، ۹٪ و ۱۰٪ از کل داده‌ها را شامل می‌شوند. در نهایت داده‌ها توکنایز می‌شوند. منظور از توکنایز، تبدیل متن به دنباله‌ای از واحدهای زبانی است. سپس در ادامه این دنباله‌ها به شناسه‌های عددی تبدیل می‌شوند تا مدل بتواند آن‌ها را پردازش کند. برای این توکن‌ها ماسک توجه نیز ایجاد می‌شود تا مدل به داده‌های مناسب، توجه بیشتر اختصاص دهد و به داده‌های نامناسب‌تر، توجه کمتری کند. این کار با استفاده از مدل پیش آموزش دیده‌ی ParsBERT[1] انجام می‌شود که نظرات را به قالب ورودی مناسب برای مدل پیشنهادی تبدیل می‌کند. این کار برای مدل‌های مبتنی بر BERT، لازم است.

۴-۲- استخراج ویژگی با استفاده از SAE

داده‌ها پس از پیش پردازش، برای استخراج ویژگی وارد مدل SAE می‌شوند. رمزگذار در این مدل شامل سه لایه متراکم است که به ترتیب دارای ۵۱۲، ۲۵۶ و ۱۲۸ نورون هستند. هر یک از این لایه‌ها از تابع فعال‌سازی ReLU[2] استفاده می‌کنند که به مدل کمک می‌کند تا روابط غیرخطی بین ویژگی‌ها را یاد بگیرد. در بخش رمزگشا، سه لایه متراکم دیگر قرار دارند که به ترتیب دارای ۲۵۶، ۵۱۲ و ۱۲۸ نورون هستند. در این بخش، از توابع فعال‌سازی ReLU و سیگموئید[3] استفاده می‌شود. تابع سیگموئید خروجی نهایی را بین ۰ و ۱ محدود می‌کند. وظیفه‌ی این بخش بازسازی داده‌های ورودی از ویژگی‌های فشرده شده‌ای است که توسط بخش رمزگذار استخراج شده‌اند. شکل ۱، معماری مدل SAE را نمایش می‌دهد. برای آموزش مدل SAE، ابتدا مدل با استفاده از بهینه‌ساز Adam و تابع هزینه‌ی خطای میانگین مربعات[4] (MSE) کامپایل می‌شود. MSE یکی از متداول‌ترین توابع هزینه در مسائل رگرسیون و بازسازی است. این تابع میزان اختلاف بین مقادیر پیش‌بینی شده توسط مدل و مقادیر واقعی را اندازه‌گیری می‌کند و هدف آن به حداقل رساندن این اختلاف است. بهینه‌ساز Adam نیز یک روش کارآمد برای بهینه‌سازی پارامترهای مدل‌های یادگیری عمیق است که ترکیبی از مزایای روش‌های AdaGrad و RMSProp را به همراه دارد.



شکل ۱ ساختار مدل SAE در مدل پیشنهادی

Adam از نرخ یادگیری تطبیقی استفاده می‌کند که به هر پارامتر مدل، یک نرخ یادگیری متفاوت اختصاص می‌دهد. این ویژگی به مدل کمک می‌کند تا به سرعت به سمت نقطه بهینه همگرا شود. در فرآیند آموزش، داده‌های آموزشی و اعتبارسنجی به مدل داده می‌شوند تا مدل بتواند پارامترهای خود را بهینه کند. در این مرحله، مدل با استفاده از داده‌های ورودی تلاش می‌کند تا ویژگی‌های مهم را استخراج کرده و داده‌های ورودی را بازسازی کند. هدف از این فرآیند، به حداقل رساندن خطای بازسازی است که توسط تابع هزینه‌ی MSE اندازه‌گیری می‌شود. پس از آموزش مدل، از بخش رمزگذار برای استخراج ویژگی‌ها از داده‌های آموزشی، اعتبارسنجی و آزمایشی استفاده می‌شود. این ویژگی‌ها به عنوان ورودی برای مدل‌های یادگیری عمیق بعدی مورد استفاده قرار می‌گیرند. با استفاده از این ویژگی‌های فشرده شده، می‌توان عملکرد مدل‌های یادگیری عمیق مانند شبکه‌های عصبی BiLSTM و CNN را بهبود بخشید. در نهایت برای استفاده از برچسب‌های داده‌ها در مدل‌های یادگیری عمیق، این برچسب‌ها به فرمت One-Hot تبدیل می‌شوند. در این فرمت، هر برچسب به یک بردار باینری تبدیل می‌شود که تنها یک عنصر آن برابر با یک و بقیه عناصر برابر با صفر هستند. این فرمت به مدل‌های یادگیری عمیق کمک می‌کند تا برچسب‌های مختلف را به درستی تشخیص داده و کلاس‌بندی انجام شود.

۴-۳- طبقه‌بندی با استفاده از مدل ترکیبی Transformer-BiLSTM-CNN

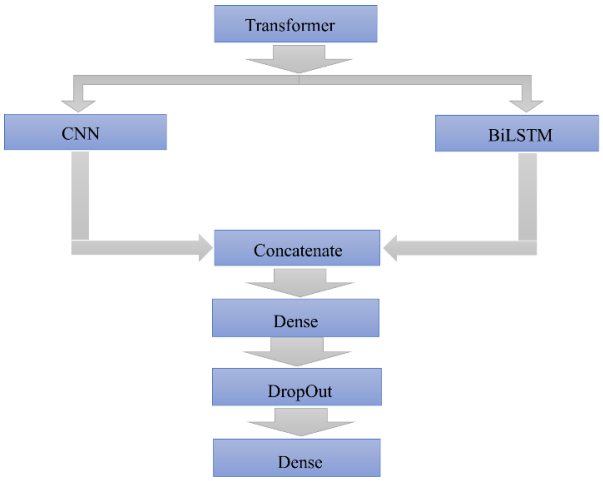
در نهایت ویژگی‌های استخراج شده در قسمت دوم، وارد مدل ترکیبی Transformer-BiLSTM-CNN شده و وظیفه‌ی طبقه‌بندی انجام می‌شود. این بخش ترکیبی از مدل‌های BiLSTM، CNN و تبدیل کننده برای طبقه‌بندی متن را پیاده‌سازی می‌کند. مدل‌های BiLSTM و CNN، به صورت موازی در کنار هم استفاده

[1] HooshvareLab/bert-fa-base-uncased
[2] rectified linear unit
[3] Sigmoid
[4] Mean Squared Error

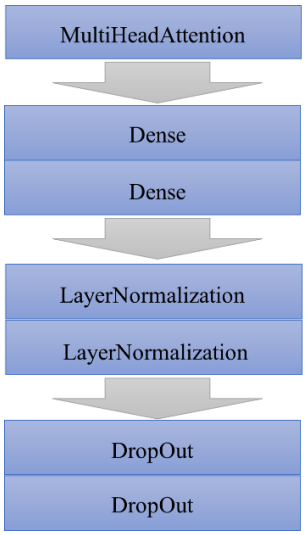
می‌شوند تا از قدرت هر کدام در استخراج ویژگی‌ها و الگوهای مختلف در داده‌های متنی بهره‌مند شویم. معماری کلی مدل ترکیبی Transformer-BiLSTM-CNN در شکل ۲ نمایش داده شده است.

در این مدل یک کلاس به نام بلوک تبدیل کننده تعریف شده است که شامل لایه‌های متفاوت است از جمله توجه چندسر که برای استخراج ویژگی‌های پیچیده از داده‌های ورودی استفاده می‌شود؛ لایه متراکم با فعالسازی ReLU و رگولاریزاسیون L2 برای پیاده‌سازی شبکه عصبی پیش‌خور؛ لایه‌ی نرمال‌سازی برای نرمال‌سازی ورودی‌ها به کار می‌رود و باعث پایداری بیشتر و سرعت‌بخشی به فرایند یادگیری می‌شود. نرمال‌سازی باعث می‌شود که میانگین خروجی‌ها صفر و انحراف معیار آن‌ها یک شود؛ لایه حذف تصادفی[1] نورون‌ها برای جلوگیری از بیش‌برازش. این بلوک تبدیل کننده به عنوان بخشی از مدل اصلی برای استخراج ویژگی‌های سطح بالا از داده‌های متنی استفاده می‌شود. شکل ۳، معماری بلوک تبدیل کننده را نمایش می‌دهد. خروجی بخش تبدیل کننده به طور موازی به BiLSTM و CNN وارد می‌شود. BiLSTM قادر به یادگیری روابط زمانی در داده‌ها است و در مدل پیشنهادی برای استخراج ویژگی‌های زمانی دو سویه استفاده می‌شود و اطلاعات متنی را در هر دو جهت زمان استخراج می‌کند. سپس از لایه GlobalMaxPooling1D استفاده می‌شود. این لایه برای کاهش ابعاد و استخراج ویژگی‌های مهم از خروجی BiLSTM به کار می‌رود و بیشینه هر ویژگی را در طول دنباله‌های زمانی انتخاب می‌کند.

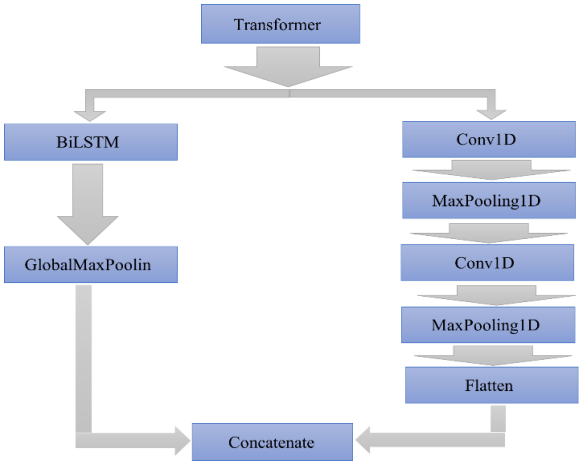
به طور موازی، CNN در مدل پیشنهادی برای استخراج ویژگی‌های محلی و فضایی استفاده می‌شود. این CNN شامل دو لایه کانولوشن یک بعدی با تعداد فیلترهای ۶۴ و اندازه هسته ۱ استفاده شده است که برای اعمال فیلترهای کانولوشن به داده‌ها استفاده می‌شوند. فیلترها به صورت هسته‌های کوچک بر روی داده‌ها اعمال می‌شوند و ویژگی‌های محلی را استخراج می‌کنند. سپس دو لایه MaxPooling1D با اندازه هسته ۱ استفاده شده است که برای کاهش ابعاد و فشرده‌سازی ویژگی‌های استخراج شده توسط لایه‌های کانولوشن به کار می‌روند. درنهایت لایه Flatten، لایه خروجی‌های چندبعدی لایه‌های قبلی را به یک بردار یک بعدی تبدیل می‌کند تا بتوان آن‌ها را به لایه‌های متراکم متصل کرد. معماری BiLSTM و CNN در شکل ۴ نمایش داده شده است.



شکل ۲ ساختار کلی مدل Transformer-BiLSTM-CNN



شکل ۳ ساختار بلوک Transformer



شکل ۴ ساختار بخشBiLSTM-CNN مدل

[1] DropOut Layer

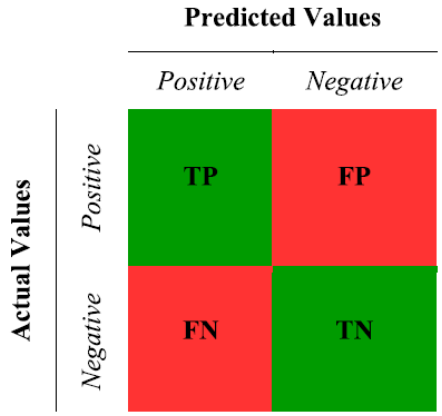
خروجی‌های BiLSTM و CNN سپس با هم ادغام شده و به یک لایه متراکم با فعال‌سازی ReLU وارد می‌شوند و در نهایت به یک لایه خروجی با فعال‌سازی سافت مکس (برای طبقه‌بندی چندکلاسه) یا سیگموئید (برای طبقه‌بندی دودویی) متصل می‌شوند. این مدل ترکیبی سپس با استفاده از بهینه‌ساز Adam و تابع هزینه متناسب با نوع مسئله کامپایل می‌شود. به طوری که بسته به نوع داده‌ها از متقاطع طبقه‌ای برای طبقه‌بندی چندکلاسه و متقاطع باینری[1] برای طبقه‌بندی دودویی استفاده می‌شود. سپس مدل با استفاده از داده‌های آموزشی و داده‌های اعتبارسنجی، آموزش داده می‌شود. این ترکیب به مدل امکان می‌دهد که از ویژگی‌های مختلف داده‌ها بهره‌برداری کند و عملکرد بهتری در وظایف طبقه‌بندی داشته باشد. در نهایت مدل آموزش‌دیده برای پیش‌بینی احتمالات تعلق هر نمونه به هر یک از کلاس‌ها استفاده می‌شود. به این ترتیب، مدل برای هر نمونه یک بردار احتمال تولید می‌کند که نشان می‌دهد هر نمونه با چه احتمالی به هر یک از کلاس‌های ممکن تعلق دارد. سپس، برای تعیین برچسب نهایی هر نمونه، بیشترین مقدار احتمال از بردار احتمال انتخاب می‌شود. این عمل به مدل امکان می‌دهد تا تصمیم‌گیری قطعی در مورد برچسب هر نمونه را بر اساس بالاترین احتمال موجود انجام دهد. این دو گام به طور مؤثر فرآیند طبقه‌بندی نمونه‌ها را با استفاده از مدل آموزش‌دیده به انجام می‌رسانند. سپس مدل با استفاده از معیارهای ارزیابی مناسب، ارزیابی می‌شود.

۵- ارزیابی مدل پیشنهادی

ما با استفاده از مدل پیشنهادی، سه مجموعه‌ی داده‌ی متفاوت را مورد بررسی قرار دادیم و نتیجه را با مدل‌های Transformer-BiLSTM-CNN [۶] و SAE-LSTM [۱۲] و یک مدل ساده CNN مقایسه کردیم. برای اینکار این مدل‌ها را مشابه مدل پیشنهادی این پژوهش توسعه دادیم و نیز داده‌های فارسی ورودی برای این مدل‌ها را به طور مشابه پیش‌پردازش کردیم تا نتایج درستی داشته باشیم.

۵-۱- معیارهای عملکرد

ماتریس سردرگمی[2]، جدولی است که عملکرد یک مدل یادگیری ماشین را بر روی یک مجموعه داده آزمون خلاصه می‌کند. این جدول تعداد پیش‌بینی‌های صحیح و نادرست مدل را نشان می‌دهد. معمولاً برای اندازه‌گیری عملکرد مدل‌های طبقه‌بندی کننده استفاده می‌شود؛ مدل‌هایی که هدف آن‌ها پیش‌بینی یک برچسب دسته‌ای برای هر نمونه ورودی است. در شکل ۵، این جدول نمایش داده شده است [۸]. مثبت واقعی[3] (TP): نتیجه آزمایشی که به درستی وجود یک وضعیت یا مشخصه را نشان می‌دهد. منفی واقعی[4] (TN): نتیجه آزمایشی که به درستی فقدان شرایط یا ویژگی را نشان می‌دهد. مثبت کاذب[5] (FP): نتیجه آزمایشی که به اشتباه نشان می‌دهد که یک شرایط یا ویژگی خاص وجود دارد. منفی کاذب[6] (FN): نتیجه آزمایشی که به اشتباه نشان می‌دهد که یک شرط یا ویژگی خاص وجود ندارد. اندازه‌گیری دقت، صحت، یادآوری و امتیاز F1 با توجه به ماتریس سردرگمی در شکل ۵ محاسبه می‌شود.



شکل ۵ ماتریس سردرگمی برای یک مسئله طبقه‌بندی باینری [۸]

دقت، نسبت تعداد پیش‌بینی‌های صحیح به تعداد کل پیش‌بینی‌ها و طبق رابطه (۱) است.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (۱)$$

صحت، برآورد کل برچسب‌های کلاس است که به طور دقیق برای هر کلاس پیش‌بینی شده است. به عبارت دیگر، نسبت تعداد پیش‌بینی‌های صحیح از یک کلاس به تعداد کل پیش‌بینی‌ها از آن کلاس اندازه‌گیری، که در رابطه (۲) بیان شده است.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (۲)$$

یادآوری، نسبت تعداد پیش‌بینی‌های صحیح از یک کلاس به تعداد کل نمونه‌های واقعی از آن کلاس که در رابطه (۳) آمده است.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (۳)$$

امتیاز F1، برای ترکیب مقادیر صحت و یادآوری در یک اندازه‌گیری استفاده می‌شود. مقدار این اندازه‌گیری بین ۰ و ۱ است و اگر

[1] Binary crossentropy
[2] confusion matrix
[3] true positive
[4] true negative
[5] false positive
[6] false negative

طبقه‌بندی کننده به درستی همه نمونه‌ها را طبقه‌بندی کند، مقدار ۱ را می‌گیرد. اندازه گیری امتیاز F1 در رابطه (۴) آورده شده است.

$$F1 = \frac{2* \text{Precision}* \text{Recall}}{\text{Precision}+\text{Recall}} \quad (۴)$$

هرچقدر مقدار امتیاز F1 به ۱ نزدیک شود؛ برای موفقیت آن کلاس بهتر است.

۵-۲- معرفی مجموعه‌ی داده

ما از نظرات کاربران فارسی زبان برای آموزش و ارزیابی مدل پیشنهادی خود بهره بردیم، این داده‌ها شامل نظر متنی کاربران به همراه امتیاز احساسی تخصیص داده شده به هر نظر است. این اطلاعات در سایت kaggle در دسترس است و نظرات کاربران مربوط به محصولات مختلف در فروشگاه‌های اینترنتی متفاوت مانند دیجیکالا[1] و کتاب‌های فروشگاه طاقچه[2] و همچنین نظرات پراکنده‌ی کاربران فارسی زبان در شبکه اجتماعی توییتر[3] است. مجموعه‌ی داده‌ی طاقچه حدود ۳۶،۰۰۰ نظر را برای مدل پیشنهادی فراهم کرده و نمره نظرات از ۰ تا ۵ دسته بندی شده بود که از ۰ تا ۲ به عنوان نظر منفی، و از ۳ تا ۵ به عنوان نظر مثبت تعیین شده‌اند. مجموعه‌ی داده‌ی دیجیکالا حدود ۲۰۰۰ نظر را فراهم کرد و نمراه نظرات از ۰ تا ۱۰۰ دسته بندی شده بود که از ۰ تا ۶۵ نظرات منفی و از ۶۵ تا ۱۰۰ نظر مثبت تعیین شدند. مجموعه‌ی داده‌ی توییتر نیز حدود ۲۰۰۰ نظر فراهم کرد که داده‌های مثبت و منفی برچسب گذاری شده بودند.

۵-۳- بحث و تفسیر نتایج

نتایج به شرح جدول‌های ۲، ۳ و ۴ است. با توجه به نتایج بدست آمده، می‌بینیم که مدل پیشنهادی در برخی از معیارها و مجموعه داده‌ها بهبود عملکرد داشته است. به عنوان مثال مدل پیشنهادی با مجموعه داده‌ی طاقچه، در مقایسه با مدل‌های دیگر، عملکرد قابل توجهی را نشان می‌دهد. دقت مدل پیشنهادی ۵۲،۰۸۸٪ است که نسبت به مدل‌های SAE-LSTM و Transformer-BiLSTM-CNN با دقت به ترتیب ۵۰،۳۶۸٪ و ۵۰٪ بهبود یافته است. علاوه بر این، صحت مدل پیشنهادی ۵۲،۱۰۵٪ است، که در مقایسه با صحت مدل SAE-LSTM با صحت ۵۰،۲۰۶٪ بهبود یافته است. مدل Transformer-BiLSTM-CNN با صحت ۷۵٪ بهترین عملکرد را در این معیار دارد، اما دقت و یادآوری پایین‌تر آن، نشان‌دهنده ناپایداری عملکرد کلی این مدل است. مدل پیشنهادی با امتیاز F1 برابر با ۵۲٪، عملکرد متعادلی را ارائه می‌دهد، که در مقایسه با مدل SAE-LSTM با امتیاز F1 برابر با ۶۵،۱۹۹٪ و مدل CNN با امتیاز F1 برابر با ۶۶،۷۱٪، نیاز به بهبود در برخی زمینه‌ها دارد اما نسبت به مدل Transformer-BiLSTM-CNN با امتیاز F1 ۳۳،۳۳٪ بهبود یافته است. در نهایت، مدل SAE-LSTM با یادآوری ۹۲،۹۵۸٪ بهترین عملکرد را در این معیار نشان می‌دهد، در حالی که مدل پیشنهادی با یادآوری ۵۲،۰۸۸٪ نیاز به بهبود در این زمینه دارد اما به نسبت مدل Transformer-BiLSTM-CNN با یادآوری ۵۰٪ بهبود یافته است. مدل پیشنهادی با مجموعه داده‌ی دیجیکالا، در مقایسه با مدل‌های دیگر، عملکرد متعادلی را نشان می‌دهد. دقت این مدل ۵۰،۲۶۱٪ است که در مقایسه با مدل SAE-LSTM با دقت ۵۲،۳۵۶٪، ضعیفتر است اما نسبت به مدل CNN با دقت ۴۷،۶۴٪ بهبود یافته است. مدل Transformer-BiLSTM-CNN با دقت ۶۰،۲۱٪ بهترین عملکرد را در این معیار دارد. صحت مدل پیشنهادی ۵۰،۲۵۹٪ است، که در مقایسه با صحت مدل SAE-LSTM ۵۴،۹۴۶٪، پایین‌تر است اما نسبت به مدل CNN با صحت ۴۷،۸۷٪ بهبود یافته است. مدل Transformer-BiLSTM-CNN با صحت ۶۰،۵۵٪ همچنان بهترین عملکرد را دارد. از نظر امتیاز F1، مدل پیشنهادی با امتیاز ۵۰،۲۵۹٪ عملکرد بهتر از مدل‌های SAE-LSTM و CNN با امتیازهای به ترتیب ۴۵،۷۹۷٪ و ۴۷،۳۷٪، دارد. مدل Transformer-BiLSTM-CNN با امتیاز F1 برابر با ۵۹،۸۳٪ بهترین عملکرد را در این زمینه دارد. در نهایت، از نظر یادآوری، مدل پیشنهادی با یادآوری ۵۰،۲۶۱٪ عملکردی بهتر از مدل CNN با یادآوری ۴۶،۸۸٪ دارد، اما نسبت به مدل‌های SAE-LSTM و Transformer-BiLSTM-CNN با یادآوری‌های به ترتیب ۵۲،۳۵۶٪ و ۶۰،۲۱٪، ضعیف‌تر عمل می‌کند. مدل پیشنهادی با مجموعه داده‌ی توییتر، در مقایسه با سایر مدل‌ها عملکرد متوسطی از خود نشان می‌دهد. دقت این مدل ۵۶،۲۸۱٪ است که در مقایسه با مدل SAE-LSTM با دقت ۴۹،۷۴۸٪، بهبود یافته است، اما نسبت به مدل Transformer-BiLSTM-CNN با دقت ۶۱،۳۱٪ و مدل CNN با دقت ۶۳،۸۲٪، ضعیفتر است. صحت مدل پیشنهادی ۵۶،۳۲۸٪ است که نسبت به صحت مدل SAE-LSTM با ۷۵٪ و مدل‌های Transformer-BiLSTM-CNN و CNN با صحت‌های به ترتیب ۶۱،۷۰٪ و ۶۶،۲۷٪، ضعیفتر است. از نظر امتیاز F1، مدل پیشنهادی با امتیاز ۵۶،۲۳۴٪ عملکرد بهتری نسبت به مدل SAE-LSTM با

[1] https://www.kaggle.com/datasets/soheiltehranipour/digikala-comments-persian-sentiment-analysis

[2] https://www.kaggle.com/datasets/armitaraz/datasettaghchecleaned/data

[3] https://www.kaggle.com/datasets/mohammadalimkh/persian-twitter-dataset-sentiment-analysis

امتیاز ۳۳,۰۵۴٪ دارد، اما نسبت به مدل‌های Transformer-BiLSTM-CNN و CNN با امتیازهای به ترتیب ۶۰,۹۳٪ و ۶۰,۴۴٪، عملکرد ضعیف‌تری دارد. در نهایت، از نظر یادآوری، مدل پیشنهادی با یادآوری ۵۶,۲۸۱٪ عملکرد بهتری نسبت به مدل SAE-LSTM و CNN با یادآوری‌های به ترتیب ۴۹,۷۴۸٪ و ۵۵,۵۶٪ دارد، اما نسبت به مدل Transformer-BiLSTM-CNN با یادآوری ۶۱,۳۱٪، ضعیف‌تر عمل می‌کند.

۵-۴- بررسی ماتریس سردرگمی مدل

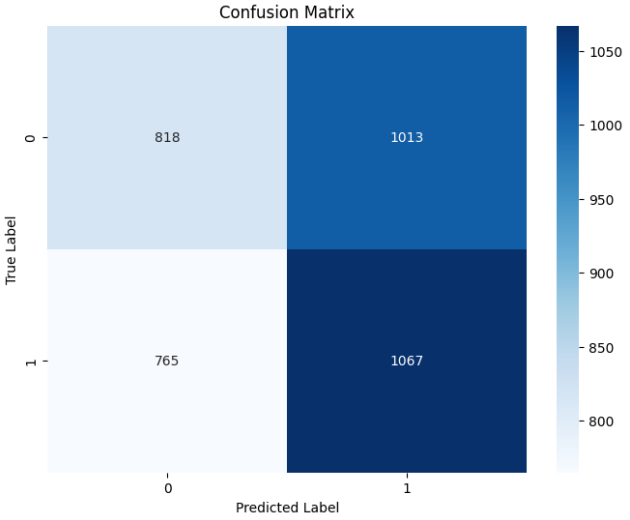
شکل ۶، ماتریس سردرگمی مدل پیشنهادی را نمایش می‌دهد. تعداد نمونه‌های مثبت واقعی که به درستی به عنوان مثبت پیش‌بینی شده‌اند، برابر با ۱۰۶۷ است. این مقدار نشان‌دهنده توانایی مدل در شناسایی صحیح نمونه‌های مثبت است. تعداد نمونه‌های منفی واقعی که به درستی به عنوان منفی پیش‌بینی شده‌اند، برابر با ۸۱۸ است. این مقدار نشان‌دهنده توانایی مدل در شناسایی صحیح نمونه‌های منفی است. تعداد نمونه‌های منفی واقعی که به اشتباه به عنوان مثبت پیش‌بینی شده‌اند، برابر با ۱۰۱۳ است. این مقدار نشان‌دهنده تعداد پیش‌بینی‌های نادرست مثبت است که مدل انجام داده است و ضعف عملکرد دقت مدل پیشنهادی را نمایان می‌کند. تعداد نمونه‌های مثبت واقعی که به اشتباه به عنوان منفی پیش‌بینی شده‌اند، برابر با ۷۶۵ است. این مقدار نشان‌دهنده تعداد پیش‌بینی‌های نادرست منفی است که مدل انجام داده است و ضعف عملکرد معیار یادآوری مدل پیشنهادی را نمایان می‌کند. به طور کلی، مدل در پیش‌بینی کلاس منفی و کلاس مثبت به صورت نابرابر عمل کرده است. مدل تعداد بیشتری از نمونه‌های مثبت را به درستی شناسایی می‌کند، اما در شناسایی نمونه‌های منفی ناتوان بوده است. این مشکلات ممکن است به دلیل پیچیدگی الگوهای داده‌ها باشد.

۵-۵- بررسی پرسش‌های پژوهش

در اینجا به بررسی پرسش‌های پژوهش می‌پردازیم.

۱) **سوال اول: چگونه با ترکیب مدل‌های یادگیری عمیق، عملکرد وظیفه تحلیل احساسات را در متن‌های فارسی بهبود دهیم؟**

۲) نتایج به دست آمده نشان می‌دهند که با ترکیب مدل‌های مختلف یادگیری عمیق مانند BiLSTM، CNN، و تبدیل‌کننده می‌توان عملکرد تحلیل احساسات را بهبود بخشید. مدل پیشنهادی در برخی موارد مانند مجموعه داده‌های طاقچه و توییتر عملکرد بهتری نسبت به مدل‌های دیگر داشته است. این بهبودها ناشی از استفاده از تکنیک‌های مختلف در پیش‌پردازش داده‌ها، استخراج ویژگی‌ها و ترکیب مدل‌های مختلف است که هر کدام از این مدل‌ها مزایای خاص خود را در بهبود دقت و صحت مدل به ارمغان می‌آورند.



شکل ۶. ماتریس سردرگمی مدل پیشنهادی

جدول ۲ نتایج روش پیشنهادی با مجموعه داده‌ی طاقچه

مدل	امتیاز F1	دقت	صحت	یادآوری
مدل پیشنهادی	٪۵۲	٪۵۲,۰۸۸	٪۵۲,۱۰۵	٪۵۲,۰۸۸
Transformer-BiLSTM-CNN	٪۳۳,۳۳	٪۵۰	٪۷۵	٪۵۰
SAE-LSTM	٪۶۵,۱۹۹	٪۵۰,۳۶۸	٪۵۰,۲۰۶	٪۹۲,۹۵۸
CNN	٪۶۶,۷۱	٪۶۹,۳۱	٪۷۲,۹۳	٪۶۱,۴۶

جدول ۳ نتایج روش پیشنهادی با مجموعه داده‌ی دیجیکالا

مدل	امتیاز F1	دقت	صحت	یادآوری
مدل پیشنهادی	٪۵۰,۲۵۹	٪۵۰,۲۶۱	٪۵۰,۲۵۹	٪۵۰,۲۶۱
Transformer-BiLSTM-CNN	٪۵۹,۸۳	٪۶۰,۲۱	٪۶۰,۵۵	٪۶۰,۲۱
SAE-LSTM	٪۴۵,۷۹۷	٪۵۲,۳۵۶	٪۵۴,۹۴۶	٪۵۲,۳۵۶
CNN	٪۴۷,۳۷	٪۴۷,۶۴	٪۴۷,۸۷	٪۴۶,۸۸

جدول ۴ نتایج روش پیشنهادی با مجموعه داده‌ی توییتر

مدل	امتیاز F1	دقت	صحت	یادآوری
مدل پیشنهادی	٪۵۶,۲۳۴	٪۵۶,۲۸۱	٪۵۶,۳۲۸	٪۵۶,۲۸۱
Transformer-BiLSTM-CNN	٪۶۰,۹۳	٪۶۱,۳۱	٪۶۱,۷۰	٪۶۱,۳۱
SAE-LSTM	٪۳۳,۰۵۴	٪۴۹,۷۴۸	٪۷۵	٪۴۹,۷۴۸
CNN	٪۶۰,۴۴	٪۶۳,۸۲	٪۶۶,۲۷	٪۵۵,۵۶

۳) سوال دوم: چگونه می‌توان با بکارگیری مدل‌های بسیار عمیق باعث بهبود استخراج ویژگی‌ها در وظیفه تحلیل احساسات شد؟

استفاده از مدل‌های بسیار عمیق مانند SAE-LSTM و Transformer-BiLSTM-CNN نشان داده که می‌توان ویژگی‌های متنی پیچیده‌تری را استخراج کرد که در نتیجه به بهبود دقت و صحت مدل‌ها کمک می‌کند. با این حال، مدل پیشنهادی نیز با استفاده از تکنیک‌های مختلف و ترکیب مدل‌ها توانسته است در برخی معیارها عملکرد بهتری داشته باشد. به عنوان مثال، مدل پیشنهادی در مجموعه داده‌های طاقچه نسبت به مدل‌های دیگر بهبود عملکرد در دقت و صحت نشان داده است. این نتایج نشان می‌دهند که استفاده از مدل‌های عمیق‌تر و تکنیک‌های پیشرفته‌تر در استخراج ویژگی‌ها می‌تواند بهبود عملکرد را به همراه داشته باشد.

۶- نتیجه‌گیری

در این پژوهش، مشکل اصلی تحلیل احساسات متنی در زبان فارسی بررسی شد که شامل چالش‌هایی نظیر پیچیدگی ساختار زبان، داده‌های کم‌منبع، و عدم تطبیق کامل مدل‌های موجود با ویژگی‌های زبانی و معنایی متن فارسی بود. این مشکلات انگیزه اصلی طراحی و ارائه یک مدل ترکیبی نوآورانه شد که با بهره‌گیری از معماری‌های پیشرفته یادگیری عمیق و تکنیک‌های پیش‌پردازش، بهبود قابل توجهی در دقت و کارایی تحلیل احساسات ارائه دهد. روش پیشنهادی با ترکیب مدل‌های Transformer-BiLSTM-CNN به منظور طبقه بندی احساسات و استفاده از SAE به منظور استخراج و کاهش ابعاد ویژگی‌ها طراحی شد. همچنین، از ParsBERT برای پیش‌پردازش متن فارسی و تولید جاسازی‌های معنایی دقیق استفاده شد. مدل پیشنهادی توانست نقاط ضعف مدل‌های پیشین، مانند ناکارآمدی در تحلیل متون طولانی و محدودیت‌های منابع محاسباتی، را کاهش دهد و عملکرد بهتری نسبت به روش‌های دیگر نظیر SAE-LSTM، Transformer-BiLSTM-CNN و CNN ارائه دهد. نتایج ارزیابی نشان داد که مدل پیشنهادی در معیارهایی نظیر دقت، F1-Score، صحت و یادآوری، عملکرد بالاتری نسبت به مدل‌های موجود دارد. همچنین، توانایی این مدل در تحلیل جملات طولانی و پیچیده فارسی، آن را برای استفاده در کاربردهای واقعی مناسب ساخته است. استفاده از SAE نیز به کاهش نیاز به منابع محاسباتی و افزایش کارایی منجر شده است. با وجود موفقیت‌های حاصل شده، این پژوهش همچنان با محدودیت‌هایی همراه است. برای مثال مدل پیشنهادی برای تحلیل احساسات چندکلاسه هنوز نیاز به بهینه‌سازی بیشتری دارد. منابع محاسباتی مورد نیاز برای آموزش مدل‌های پیش‌آموزش‌دیده نظیر ParsBERT همچنان چالشی برای محیط‌های محدود به منابع است. داده‌های موجود برای زبان فارسی در مقایسه با زبان‌های دیگر همچنان کم‌تنوع هستند که می‌تواند عملکرد مدل را در کاربردهای خاص محدود کند. استفاده از مدل SAE برای استخراج ویژگی در حوزه تحلیل احساسات متنی بسیار جدید است و منابع محدودی در این زمینه موجود است.

در آینده، می‌توان پژوهش را به چندین جهت گسترش داد:

۱) می‌توان برای بهبود عملکرد SAE، تعداد لایه‌های رمزگذار و رمزگشا را کاهش و یا افزایش داد. این لایه‌های رمزگذار، داده‌های متنی را رمزگذاری می‌کند و سپس با رمزگشایی از آن، ویژگی‌های مهم را استخراج کرد اما بسته به نوع داده‌ها و ویژگی‌های مهم، می‌توان تعداد لایه‌های رمزگذار و رمزگشا را تغییر داد تا متناسب با داده‌های ورودی، عملکرد بهتری در زمینه‌ی استخراج ویژگی داشته باشد.

۲) می‌توان با استفاده از تکنیک‌هایی مانند خراش دادن مرورگر[1] و خزیدن در مرورگر[2] داده‌های متنوع و دست اول بدست آورد و با استفاده از تکنیک‌های داده‌کاوی، داده‌ها را آماده‌ی استفاده در مدل پیشنهادی کرد.

۳) این پژوهش برای بررسی و بهبود مدل‌های مشابه مورد استفاده قرار گرفت. این مدل پیشنهادی می‌تواند در یک پروژه‌ی کاربردی مورد استفاده قرار گیرد و وظیفه‌ی افکارسنجی را از کاربران فارسی زبان شبکه‌های اجتماعی، حول محور موضوعات اجتماعی، سیاسی و اقتصادی انجام دهد؛ تا این مدل، یک محک عملی هم ببینند.

مراجع

[1] S. Tam, R. BenSaid and Ö. Ö. Tanriöver, "A ConvBiLSTM Deep Learning Model-Based Approach for Twitter Sentiment Classification," IEEE Access, vol. 9, Mar. 2021

[2] F. Huang, X. Li, C. Yuan, S. Zhang, J. Zhang and S. Qiao, "Attention-Emotion-Enhanced Convolutional LSTM for Sentiment Analysis," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, Sept. 2022

[3] L. Zhang, S. Wang and B. Liu, "Deep learning for sentiment analysis: A survey," wires data mining and knowledge discovery, vol. 8, Aug. 2018

[1] Web Scraping

[2] Web Crawler

[4] M. Dong, Y. Li, X. Tang, J. Xu, S. Bi and Y. Cai, " Variable Convolution and Pooling Convolutional Neural Network for Text Sentiment Classification," IEEE Access, vol. 8, jan. 2020

[5] L. Yue, W. Chen, X. Li, W. Zuo and M. Yin, "A survey of sentiment analysis in social media," Knowledge and Information Systems, vol. 60, July. 2018

[6] M. Shukla and A. Kumar, "An Experimental Analysis of Deep Neural Network Based Classifiers for Sentiment Analysis Task," IEEE Access, vol. 11, Apr. 2023

[7] M. Birjali, M. Kasri and A. B. Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," Knowledge-Based Systems, vol. 226, Aug. 2021

[8] M. U. Salur and I. Aydin, "A Novel Hybrid Deep Learning Model for Sentiment Classification," IEEE Access, vol. 8, Mar. 2020

[9] D. W. O. Otter, J. R. Medina and J. K. Kalita, " A Survey of the Usages of Deep Learning for Natural Language Processing," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, Apr. 2020

[10] F. A. Acheampong, H. N. Mensah and W. Chen, "Transformer models for text‐based emotion detection: a review of BERT‐based approaches," Artificial Intelligence Review, vol. 54, Feb. 2021

[11] K. Berahmand, F. Daneshfar, E. S. Salehi, Y. Li and Y. Xu, " Autoencoders and their applications in machine learning: a survey," Artificial Intelligence Review, vol. 57, Feb. 2024

[12] I. Kanwal, F. Wahid, S. Ali, A. UrRehman, A. Alkhayyat and A. Al-Radaei, "Sentiment Analysis Using Hybrid Model of Stacked Auto-Encoder-Based Feature Extraction and Long Short Term Memory-Based Classification Approach," IEEE Access, vol. 11, Sept. 2023

[13] G. M. Shafiq, T. Hamza, M. F. Alrahmawy and R. El-Deeb, "Enhancing Arabic Aspect-Based Sentiment Analysis Using End-to-End Model," IEEE Access, vol. 11, Dec. 2023

[14] M. Aljebreen, B. Alabduallah, M. M. Asiri, A. S. Salama, M. Assiri and S. S. Ibrahim, "Moth Flame Optimization With Hybrid Deep Learning Based Sentiment Classification Toward ChatGPT on Twitter," IEEE Access, vol. 11, Sept. 2023

[15] I. Carvalho, H. G. Oliveira and C. Silva, "The Importance of Context for Sentiment Analysis in Dialogues," IEEE Access, vol. 11, Aug. 2023

[16] K. L. Tan, C. P. Lee, K. M. Lim and K. S. M. Anbananthen, "Sentiment Analysis With Ensemble Hybrid Deep Learning Model," IEEE Access, vol. 10, Sept. 2022

[17] K. L. Tan, C. P. Lee, K. S. M. Anbananthen and K. M. Lim, "RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network," *IEEE Access*, vol. 10, Feb. 2022

[18] S. Seo, C. Kim, H. Kim, K. Mo, P. Kang, "Comparative Study of Deep Learning-Based Sentiment Classification," *IEEE Access*, vol. 8, Jan. 2020

[19] N. Jin, J. Wu, X. Ma, K. Yan and Y. Mo, "Multi-Task Learning Model Based on Multi-Scale CNN and LSTM for Sentiment Classification," *IEEE Access*, vol. 8, Apr. 2020

[20] R. Zeng, H. Liu, S. Peng, L. Cao, A. Yang, C. Zong and G. Zhou, "CNN-Based Broad Learning for Cross-Domain Emotion Classification," Tsinghua Science and Technology, vol. 28, Apr. 2023

[21] I. K. S. Al-Tameemi, M. R. F. Derakhshi, S. Pashazadeh and M. AssadPour, "Interpretable Multimodal Sentiment Classification Using Deep Multi-View Attentive Network of Image and Text Data," *IEEE Access*, vol. 11, Aug. 2023

[22] X. L. Leng, X. A. Miao, and T. Liu, "Using recurrent neural network structure with Enhanced Multi-Head Self-Attention for sentiment analysis," Multimedia Tools and Applications, vol. 80, Jan. 2021

[23] M. R. R. Rana, A. Nawaz, T. Ali, A. S. Alattas and D. S.AbdElminaam, "Sentiment Analysis of Product Reviews Using Transformer Enhanced 1D-CNN and BiLSTM," Cybernetics and Information Technologies, vol. 24, Sept. 2024

[24] B. Yosra and M. Hakim, "Enhancing Twitter Sentiment Analysis Using Hybrid Transformer and Sequence Models," Japan Journal of Research, vol. 6, Oct. 2024.