Knowledge-Based Systems 198 (2020) 105949

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques

Mohammad Ehsan Basiri [a,*], Moloud Abdar [b], Mehmet Akif Cifci [c], Shahla Nemati [a], U. Rajendra Acharya [d,e,f]

[a] Department of Computer Engineering, Shahrekord University, Shahrekord 64165478, Iran
[b] Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia
[c] Department of Computer Engineering, Istanbul Aydin University, Istanbul No:38, 34295, Turkey
[d] Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore
[e] Department of Bioinformatics and Medical Engineering, Asia University, Taichung, Taiwan
[f] International Research Organization for Advanced Science and Technology (IROAST) Kumamoto University, Kumamoto, Japan

ARTICLE INFO

Article history:
Received 23 February 2020
Received in revised form 12 April 2020
Accepted 19 April 2020
Available online 22 April 2020

Keywords:
Machine learning
Deep learning
Sentiment analysis
Information fusion
3-way decisions
Drug review

ABSTRACT

Nowadays, the development of new computer-based technologies has led to rapid increase in the volume of user-generated textual content on the website. Patient-written medical and health-care reviews are among the most valuable and useful textual content on social media which have not been studied extensively by researchers in the fields of natural language processing (NLP) and data mining. These reviews offer insights into the interaction of patients with doctors, treatment, and their satisfaction or frustration with the delivery of healthcare services. In this study, we propose two deep fusion models based on three-way decision theory to analyze the drug reviews. The first fusion model, 3-way fusion of one deep model with a traditional learning algorithm (3W1DT) developed using a deep learning method as a primary classifier and a traditional learning method as the secondary method that is used when the confidence of the deep method during classification of test samples is low. In the second proposed deep fusion model, 3-way fusion of three deep models with a traditional model (3W3DT), three deep and one traditional models are trained on the entire training data and each classifies the test sample individually. Then, the most confident classifier is selected to classify the test drug review. Our results on the reviews based on Drugs.com dataset show that both proposed 3W1DT and 3W3DT methods outperformed the traditional and deep learning methods by 4% and the 3W3DT outperformed 3W1DT by 2% in terms of accuracy and F1-measure.

1. Introduction

Advancement in technology has made sentiment analysis (SA) an increasingly popular field of study in recent years focusing on the treatment of subjectivity in textual content [1]. The rapid growth of social media and advances in natural language processing (NLP) methods have paved the way for SA techniques to mine user-generated data in various fields such as tourism, marketing, and politics [2]. Nevertheless, medical, and health-care domains have been less investigated by researchers in the SA field.

With the emergence of web and rapid growth of social media, it is now a common practice for patients to share their views on treatment and drugs on social media platforms. This produces significant amount of unstructured textual data which may be utilized in important applications including finding adverse drug reactions (ADRs), assisting in diagnosis and treatment choices, recommending personalized therapy options, and finding serendipitous drug usage [3]. These applications traditionally utilize structured data which is difficult to generate and limited in quantity. On the other hand, user-generated reviews do not have such limitations making them a promising alternative for structured data.

In this study, patient-written drug reviews are analyzed to understand users' view on various drugs whose safety traditionally depend on few clinical trials and specific test protocols observed during the limited test duration [24]. Specifically, SA is applied on the drug reviews to find whether a review is positive, negative, or neutral. The SA of drug reviews is an important task because after consuming drugs, users share their experiences and preferences anonymously, providing a great deal of information to be examined to make accurate decision about public health and

* Corresponding author.
E-mail addresses: basiri@sku.ac.ir (M.E. Basiri), m.abdar1987@gmail.com (M. Abdar), wwwakif@msn.com (M.A. Cifci), s.nemati@sku.ac.ir (S. Nemati), aru@np.edu.sg (U.R. Acharya).

https://doi.org/10.1016/j.knosys.2020.105949
0950-7051/

Table 1
Comparison of recent studies conducted using sentiment analysis.

Study	Type	Description
Basiri and Kabiri [4]	Lexicon-based	Compared lexicon-based methods for SA in the Persian language
Duwairi et al. [5]		Proposed lexicon-based method for Arabic SA
Zhang et al. [6]		Proposed topic-specific Cinease lexicon for SA
Plaza-del-Arco et al. [7]		Improved lexicon-based emotion recognition in Spanish
Asghar et al. [8]		Improved lexicon-based SA using rule-based
Vashishtha and Susan [9]		Improved lexicon-based SA using fuzzy rule-based
Basiri et al. [10]		Improved lexicon-based SA using target detection
Basiri and Kabiri [11]		Improved Persian lexicon-based SA using lexicon refining
Basiri and Kabiri [12]	Hybrid	Improved Persian lexicon-based SA using machine learning
Teng et al. [13]		Combining deep learning with lexicon-based SA
Hew et al. [14]	ML-based	Used gradient boosting trees for SA
Rintyarna et al. [15]		Improved SA using domain sensitive features
Wu et al. [16]		Improving semi-supervised SA using with variational autoencoder
Han et al. [17]	Multi-modal	Improving SA using dynamic threshold and multi-classifiers
Poria et al. [18]		Combined text, audio, and image for SA
Zhao et al. [19]		Combined image and text for SA
Majumder et al. [20]		Used hierarchical fusion with context modeling for SA
Nemati et al. [21]		Combined audio, video, and text for emotion recognition

Table 2
Comparison of recent studies conducted using sentiment analysis on drug reviews.

Study	Type	Description
Ebrahimi et al. [22]	Rule-based + ML	Specified adverse drugs' side effects.
Chew and Khoo [23]	Analytical	Compared drug-related information on social media and authoritative health web sites.
Asghar et al. [8]	Lexico-based SA	Proposed an enhanced lexicon-based method to find adverse drug reactions
Gräßer et al. [24]	ML	Addressed the problem of cross-data and cross-domain learning on drug reviews
Jiménez-Zafra et al. [25]		Introduced two datasets of drugs- and doctors-related reviews from Spanish web sites
Ru et al. [3]	Deep	Identified serendipitous drug usage using a context-aware deep neural network
Liu et al. [26]		Proposed a new position-aware word embedding and generate an enhanced medical lexicon
Zhang et al. [27]		Proposed a weakly supervised neural model for drug review sentiment analysis.
Han et al. [28]		Introduced a new aspect-level drug review dataset

Table 3
Comparison of recent studies in deep models for sentiment analysis.

Study	Type	Description
Do et al. [29]	Aspect-level SA	Reviewed the aspect-based sentiment analysis by researchers and various deep learning strategies
Young et al. [30]	NLP	Summarized and compared various models in SA
Zhang et al. [31]	SA	Surveyed various applications related to SA using deep learning
Liu et al. [32]	Document-level SA	Combined 2D CNN and attention-based bidirectional GRU for sentiment classification
Kumar et al. [33]	Multimodal SA	Utilized the real time multimodal data for fine-grained prediction of sentiments
Fa et al. [34]	Document-level	Presented an adverse drug event detection and extraction mechanism from the open data
Park et al. [35]	Aspect-level	Devised a deep learning strategy based on holistic recurrent attention on target-dependent memories
Minaee et al. [36]	Document-level	Devised a model for sentiment analysis based on ensemble of CNN and LSTM.
Ma et al. [37]	Aspect-level	Devised a hierarchical attention technique with LSTM comprising of sentence-level and target-level attention.

drug safety [38]. Sentiment analysis results are useful for patients, drug manufacturers and clinicians to gain valuable insights about the potential risks of adverse drug reactions.

The existing studies on drug review using SA can be categorized into *two* main groups: traditional feature-based machine learning techniques and deep learning methods [28]. In the first category, the performance of the system relies heavily on manually designed features extracted from the textual content of reviews. Deep learning methods, on the other hand, apply different neural networks such as convolutional neural network (CNN) or long short-term memory (LSTM) on the dense vector representations of textual content [35]. Although the deep learning methods usually outperform the traditional methods, their interpretability and time complexity are considered as their drawbacks [34]. Another difference in the two methods is their ability to classify different types of test samples [39]. This inspired us to propose a fusion method which will benefit from both deep and traditional learning methods for drug review classification.

In order to use the advantages of traditional and deep learning methods effectively, this study used the theory of three-way decisions that is a new interpretation of rules in rough set theory [40].

Table 4
Thresholds used to convert ratings to labels and distribution of three classes of drug reviews.

Rating	Label	Class name	Percentage
≤ 4	-1	Negative	25%
$4 < r < 7$	0	Neutral	9%
≥ 7	$+1$	Positive	66%

In this theory, the test sample space is divided into *three* regions: accept, reject, and delay [41]. The first two regions correspond to high-confidence classification results obtained for positive and negative classes whereas the delay region corresponds to low-confidence results. In the current study, following the approach presented in [24], the problem is to classify test drug reviews into positive, negative, and neutral categories. Therefore, we adopted three regions in the original three-way decisions into four regions, one corresponding to the low-confidence decisions and the remaining high-confidence regions for three possible categories.

Different from previous studies, we propose *two* fusion models named 3-way fusion of one deep model with a traditional

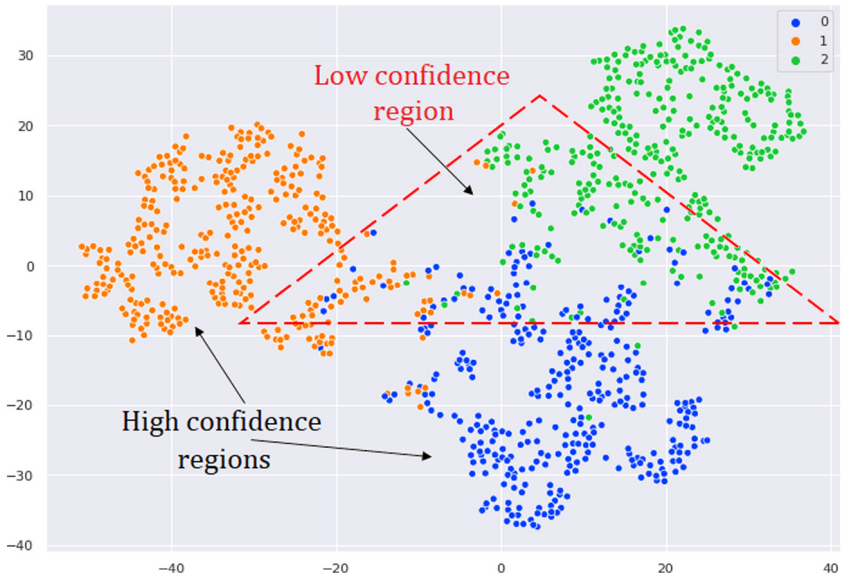

Fig. 1. Distribution of three classes of drug reviews based on the classifier confidence.

model (3W1DT) and 3-way fusion of three deep models with a traditional model (3W3DT) to utilize both deep and traditional feature-based learning methods. To specify different regions described in 3-way decisions based in the confidence level of the classifier in classification of a test sample, we propose a mechanism based on the values obtained at the output layer of deep learning methods and classification results of traditional feature-based learning methods.

In summary, the main *three* contributions of this study are as follows:

- Proposed *two* fusion models for utilizing deep learning and traditional feature-based machine learning methods for SA of drug reviews.
- Exploited 3-way decisions theory to effectively find accurate label of samples.
- Compared our proposed models with four traditional supervised machine learning and seven deep learning techniques using Drugs.com dataset.

2. Related work

2.1. Sentiment classification

Sentiment classification has become a hot topic of research in NLP community, and numerous classification methods have been proposed in recent years [2]. In this section, we compared few recent studies in sentiment analysis (SA). We categorized these studies based on the core classification methods employed and the data type applied to the method as shown in Table 1. Comprehensive surveys on the challenges and applications of SA are discussed in [42] and [43].

2.2. Drug review classification

Drug reviews are important due to several reasons; First, patients will know the results of using drugs by others like them and such information is usually cannot be easily found from their friends and family. Second, it has been shown that online community information and others' experience have positive effect on some kind of diseases [44]. Third, there exists several important applications such as identification of adverse drug effects and drug recommendation which may benefit from the results of SA of drug reviews. Nevertheless, drug reviews have been less analyzed using SA by researches as compared to other types of reviews such as customer feedback, financial, or political reviews.

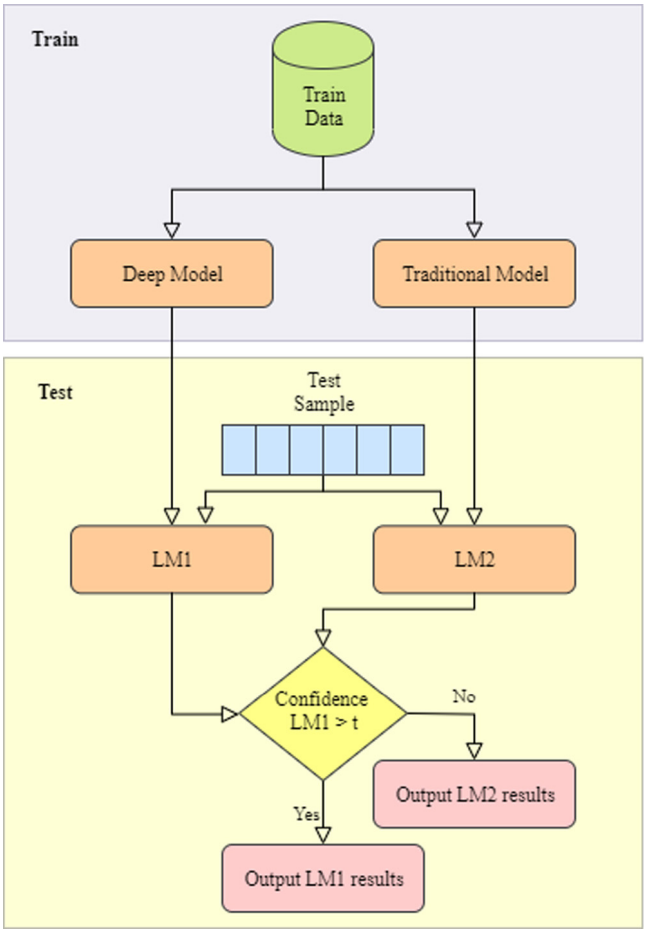

Fig. 2. Structure of 3W1DT framework obtained by the fusion of deep model (LM1) and traditional model (LM2).

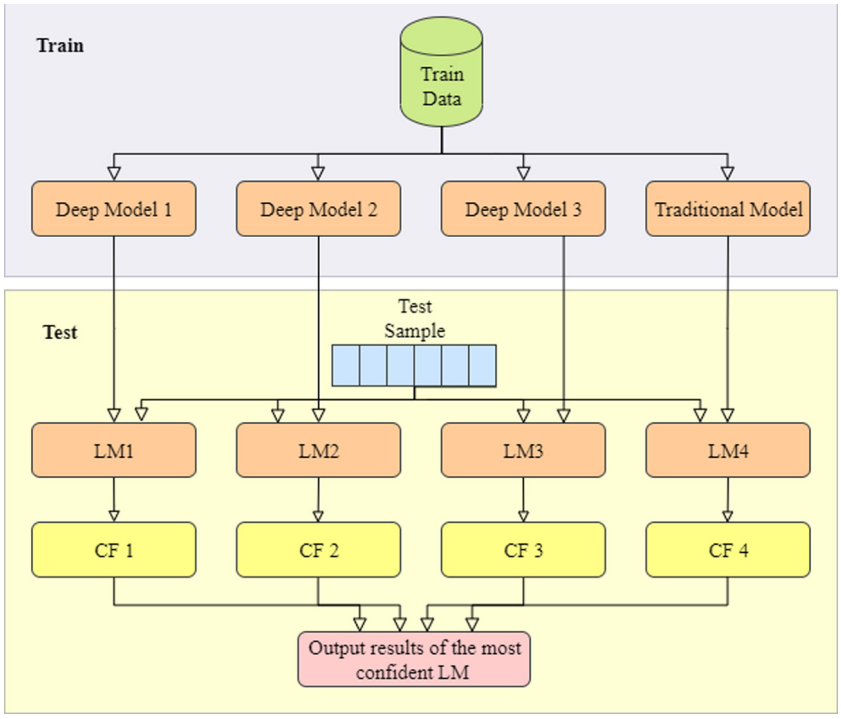

Fig. 3. Structure of 3W3DT framework obtained by the fusion of three deep model (LM1, LM2, and LM3) and a traditional model (LM2).

In Table 2, a brief review of few recent related works on drug reviews using sentiment classification is presented.

2.3. Deep learning for sentiment analysis

Deep learning is emerging as a vital tool in the ML area as it does not require feature engineering. For sentiment analysis, deep models have shown promising results. In Table 3, some of most related studies are compared.

3. Methodology

In this study, *two* deep fusion models are proposed for the classification of drug reviews based on the three-way decision theory. In the original three-way decision theory, the results of classification are divided into *three* regions: accept, reject and abstain decisions [41]. This separation is suitable for binary decisions where the output of the classifier is either positive or negative (accept and reject decisions). In such cases the abstain decision corresponds to those outputs for which the classifier confidence is lower than a pre-defined threshold. In the current study, there are *three* classes of drug reviews, namely, positive, negative, and neutral. Therefore, the decision space is divided into four regions in which *three* parts corresponds to the *three* classifier decisions and the last part belongs to the boundary or uncertain decisions. It is shown in Fig. 1.

As shown in Fig. 1, when the test sample is not in the low-confidence section, the confidence level of classifier is high and hence, we can rely on the decision made by the classifier. On the other hand, when the test sample is in the boundary region, the classifier's decision is not reliable and hence, we need more information to classify the sample correctly.

3.1. Proposed frameworks

To address the above-mentioned problem, *two* deep fusion frameworks are proposed. In the first proposed framework, 3-way fusion of one deep model with a traditional model (3W1DT), a deep learning model is used as the base classifier (BC) to classify the test samples in the high-confidence region. A traditional machine learning classifier such as multinomial Naïve Bayes or decision tree is used as secondary classifier (SC) for the samples in the boundary region. The overall view of the proposed 3W1DT is shown in Fig. 2.

To decide whether to use the SC or rely on the BC decision, we compared the outputs of the deep classifier (BC) with a

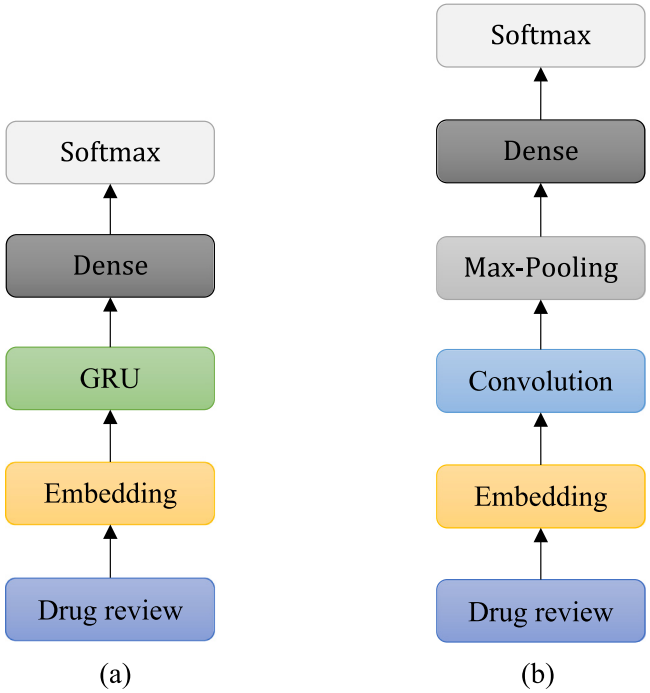

Fig. 4. Deep models used in the 3W3DT framework: (a) GRU and (b) CNN model.

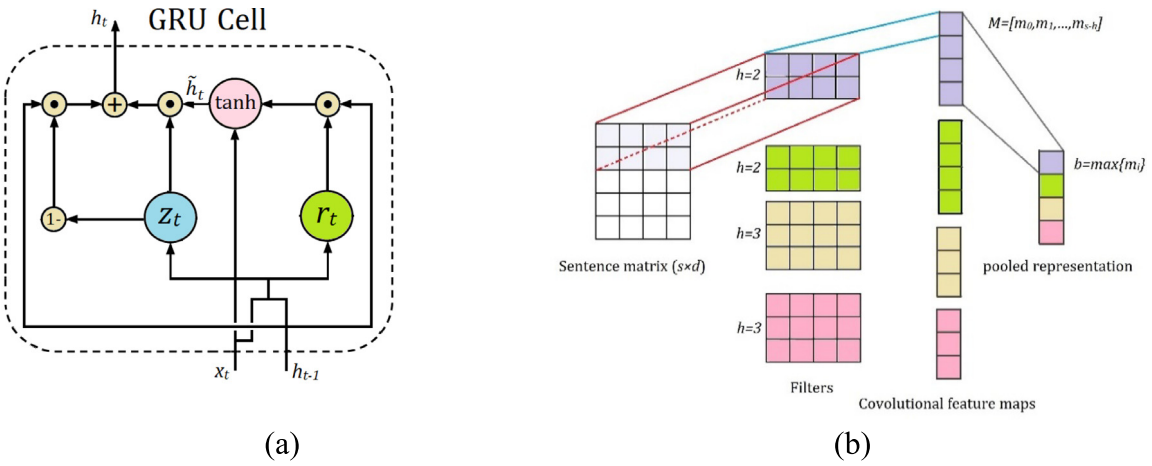

Fig. 5. The details of (a) the GRU and (b) the CNN cells.

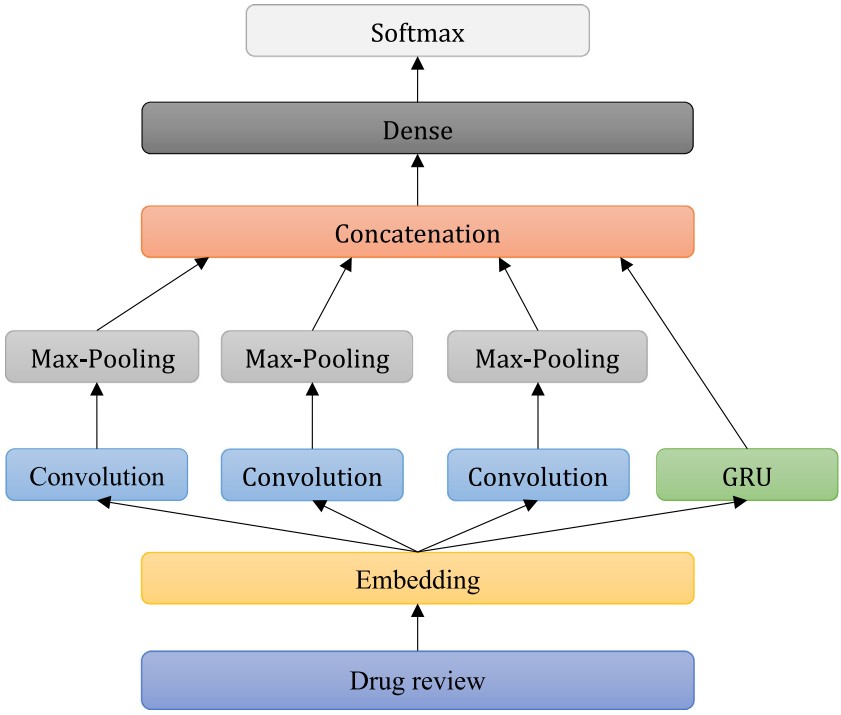

Fig. 6. The overall view of 3CRNN model used in the 3W3DT framework.

threshold t found empirically. The output layer of the deep classifier consists of *three* softmax neurons (see Section 3.1.3), each corresponding to the classification confidence of one of possible three classes [24] (i.e., positive, negative, and neutral classes). The greater the distance between the results of the three output neurons, the greater will be the certainty of the answer. In other words, the confidence of the classifier may be calculated based on the following value:

$$Cf\ (BC) = |o_1 - o_2 - o_3| \tag{1}$$

where, o_1, o_2, and o_3 are the outputs of Softmax neurons in the output layer of the deep model.

In the second proposed deep fusion model for drug review classification, 3-way fusion of three deep models with a traditional model (3W3DT), three deep and one traditional models are trained on the entire training data and each classifies the test sample individually. The outputs of these models are in the form of 3W1DT model (i.e., they are three probabilities each corresponds to the confidence in one of three possible classes). Therefore, for each model, the confidence value can be calculated using a formula similar to Eq. (1). These confidence values are CF_1, CF_2, CF_3, and CF_4 as shown in Fig. 3.

After calculating CF values for each test sample, the output of the 3W3DT framework is given by:

$$SM = \underset{1\le i\le 4}{\operatorname{argmax}}\, \mathrm{CF}(LM_i) \tag{2}$$

where, LM_i is the ith learning model, SM is the selected model for generating the final output, and CF(LM) is calculated according to Eq. (1) [41].

In the 3W3DT framework, three deep learning models namely, gated recurrent unit (GRU), CNN, and three-way convolutional recurrent neural network (3CRNN) are used. These models are shown in Figs. 4 and 5.

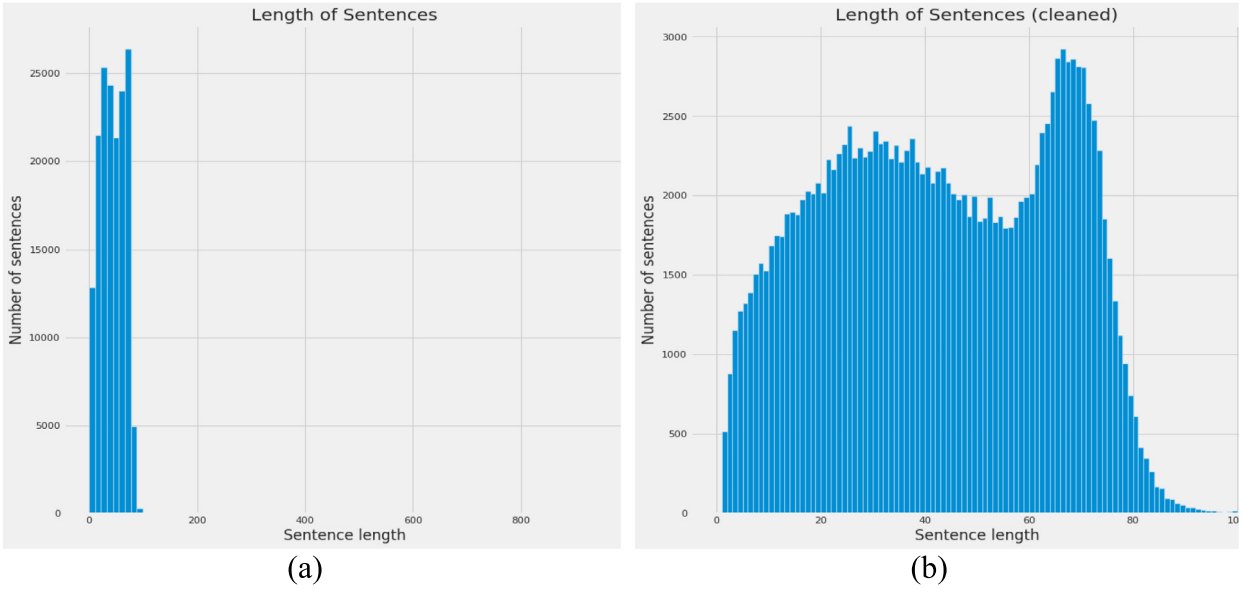

Fig. 7. Plot of number of sentences versus sentence length: (a) before (b) after removing outlier-length reviews.

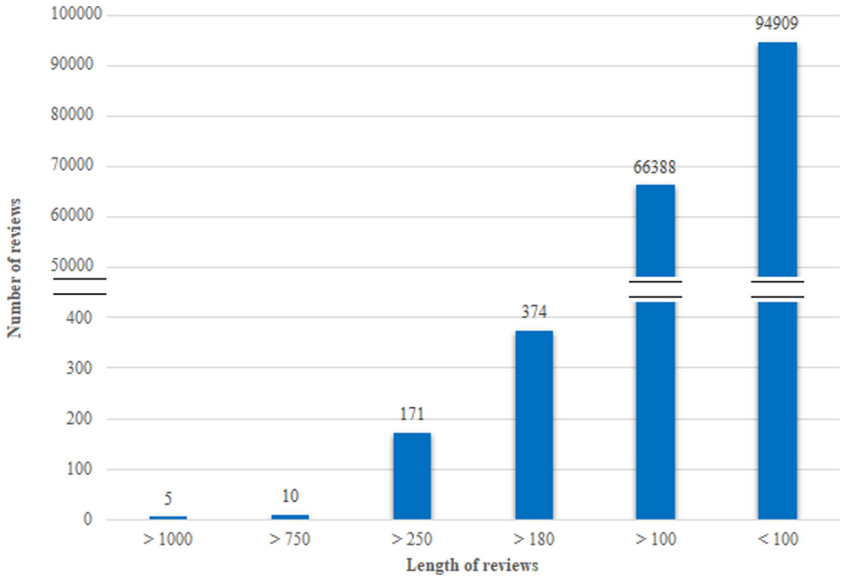

Fig. 8. Plot of number of reviews versus length of reviews.

3.1.1. GRU model

In the GRU model, the first layer after the input layer is an embedding layer. This layer is used to map the input feature vector which is a sequence of word indexes to dense vectors of fixed size. The next layer is a recurrent layer of type gated recurrent unit (GRU) (Fig. 5(a)). The GRU is a simpler variant of long short-term memory (LSTM) which is a special type of recurrent neural network (RNN) [32]. The GRU layer is usually used in text-related problems where the sequence is more important than the individual tokens which is the case in drug review classification.

Similar to LSTM cells, GRU is designed to handle the vanishing/exploding problem of RNNs. This problem occurs when the gradient becomes very small. In such cases, no real learning is performed because the parameter update becomes meaningless. A GRU cell has only two gates, an update gate r combining forget and input gates that are also used in LSTMS, and a reset gate z. The update and reset mechanisms are produced via the following functions:

$$r_t = \delta(W_r h_{t-1} + U_r x_t + b_r) \tag{3}$$

$$z_t = \delta(W_z h_{t-1} + U_z x_t + b_z) \tag{4}$$

where, δ is the logistic sigmoid function, U and W show the weight matrices of gates or cell for input x_t and hidden state h_t, and b denotes the bias vector. The reset gate is used to decide when the previous hidden state should be ignored. The update gate is responsible for the amount of input information to the current state [32]. The hidden state is computed via the following functions:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{5}$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}_t} (h_{t-1} \odot r_t) + U_{\tilde{h}_t} x_t) \tag{6}$$

Table 5
Comparison of the results obtained using seven different algorithms for the first experiment.

Method	Classes	Precision	Recall	F1-score
NB	Positive	0.7995	0.8584	0.8279
	Natural	0.9590	0.3730	0.5371
	Negative	0.8961	0.9455	0.9201
	weighted avg	**0.8775**	**0.8722**	**0.8626**
DT	Positive	0.5106	0.2196	0.3071
	Natural	0.0000	0.0000	0.0000
	Negative	0.6926	0.9373	0.7966
	weighted avg	0.5847	0.6729	0.6022
RF	Positive	0.0000	0.0000	0.0000
	Natural	0.0000	0.0000	0.0000
	Negative	0.6592	1.0000	0.7946
	weighted avg	0.4345	0.6592	0.5237
KNN	Positive	0.4663	0.4082	0.4353
	Natural	0.2679	0.0843	0.1282
	Negative	0.7372	0.8410	0.7857
	weighted avg	0.6270	0.6644	0.6387
GRU	Positive	0.6974	0.7208	0.7089
	Natural	0.3279	0.0460	0.0806
	Negative	0.8354	0.9226	0.8768
	weighted avg	0.7552	0.7932	0.7631
CNN	Positive	0.8049	0.7932	0.7990
	Natural	0.5896	0.4164	0.4881
	Negative	0.8960	0.9368	0.9159
	weighted avg	0.8456	0.8540	0.8481
3CRNN	Positive	0.8081	0.8039	0.8060
	Natural	0.6001	0.4347	0.5041
	Negative	0.9034	0.9391	0.9209
	weighted avg	0.8522	0.8598	0.8546

The output of the GRU layer is sent to the output layer via a dense layer. The output layer consists of three softmax cells, each corresponding to one of three possible classes of drug reviews.

3.1.2. CNN model

Similar to the GRU model, the first layer on the top of the input layer is an embedding layer. The next layer in this model is the convolution layer which is used for local feature extraction [45]. The convolution operation is performed on the input features via linear filters. To apply CNN to sentence S with s words, first, an embedding vector of size e is created. Filter F of size $e \times h$ is then repeatedly applied to the sub-matrices of the input matrix. This, produces a feature map $M = [m_0, m_1, \ldots, m_s - h]$ as follows:

$$m_i = F \cdot S_{i:i+h-1} \tag{7}$$

where, $i = 0, 1, \ldots, s - h$ and $S_{i:j}$ is a sub-matrix of S from row i to j. It is a standard procedure to shrink the size of feature map by feeding them to the pooling or sub-sample layer. In the CNN model used in the current study, global max-pooling is used to consolidate the output from the convolutional layer. Max-pooling is a popular pooling technique which selects the most significant feature b of the feature map as follows:

$$b = \max_{0 \le i \le s-h} \left(m_j\right) \tag{8}$$

The outputs of pooling layer are concatenated to form a pooled feature vector which is then used as the input to fully connected network (see Fig. 5(b)). Finally, the output layer in the CNN models consists of three Softmax cells similar to GRU model described earlier.

3.1.3. 3CRNN model

The first *two* layers of 3CRNN [45] model are similar to GRU and CNN. On the top of these layers, 3CRNN model uses *three* parallel convolution neural networks which apply different kernel sizes at the output of embedding layer. This creates a multi-channel CNN which reads text with different n-gram sizes (word groups). A GRU layer is also applied on the embedding layer in parallel with the multichannel CNN model. This layer is added to equip the 3CRNN model with the advantages of the GRU model in processing long dependencies in drug reviews. The output of these *four* layers are concatenated and passed to the dense layer as shown in Fig. 6.

Table 6
Comparison of the results obtained using the first fusion model for the second experiment.

Method	Classes	Precision	Recall	F1-score
GRU-NB	Positive	0.8285	0.7983	0.8131
	Natural	0.9580	0.2454	0.3907
	Negative	0.8652	0.9649	0.9124
	weighted avg	0.8643	0.8585	0.8406
CNN-NB	Positive	0.8506	0.8216	0.8359
	Natural	0.7785	0.4026	0.5307
	Negative	0.8890	0.9591	0.9227
	weighted avg	0.8694	0.8746	0.8657
3CRNN-NB	Positive	0.8549	0.8280	0.8412
	Natural	0.7935	0.4090	0.5398
	Negative	0.8925	0.9621	0.9260
	weighted avg	**0.8742**	**0.8788**	**0.8700**
GRU-DT	Positive	0.7738	0.4132	0.5387
	Natural	0.3462	0.0037	0.0074
	Negative	0.7405	0.9717	0.8405
	weighted avg	0.7134	0.7446	0.6899
CNN-DT	Positive	0.8551	0.6066	0.7097
	Natural	0.6526	0.2116	0.3196
	Negative	0.8050	0.9683	0.8792
	weighted avg	0.8039	0.8095	0.7864
3CRNN-DT	Positive	0.8546	0.5643	0.6798
	Natural	0.6568	0.1986	0.3050
	Negative	0.7939	0.9721	0.8740
	weighted avg	0.7968	0.8002	0.7741
GRU-RF	Positive	0.8910	0.3193	0.4702
	Natural	0.3462	0.0037	0.0074
	Negative	0.7186	0.9910	0.8331
	weighted avg	0.7284	0.7337	0.6678
CNN-RF	Positive	0.9076	0.5554	0.6891
	Natural	0.6526	0.2116	0.3196
	Negative	0.7879	0.9769	0.8723
	weighted avg	0.8058	0.8023	0.7767
3CRNN-RF	Positive	0.9221	0.5069	0.6542
	Natural	0.6568	0.1986	0.3050
	Negative	0.7751	0.9817	0.8663
	weighted avg	0.8014	0.7922	0.7626
GRU-KNN	Positive	0.7269	0.5424	0.6212
	Natural	0.3815	0.0590	0.1022
	Negative	0.7825	0.9483	0.8575
	weighted avg	0.7326	0.7665	0.7303
CNN-KNN	Positive	0.8179	0.6753	0.7398
	Natural	0.6117	0.2528	0.3578
	Negative	0.8356	0.9579	0.8926
	weighted avg	0.8111	0.8236	0.8062
3CRNN-KNN	Positive	0.8155	0.6507	0.7239
	Natural	0.6106	0.2458	0.3505
	Negative	0.8298	0.9611	0.8906
	weighted avg	0.8065	0.8190	0.8003

3.2. Novelty of the proposed framework

As discussed in the earlier section, *two* methods are proposed in this study to classify sentiment expressed in drugs reviews. The main novelty of these methods is the combination of traditional and deep models using 3-way decision theory. Consequently, these methods take into account the uncertainty existed in the

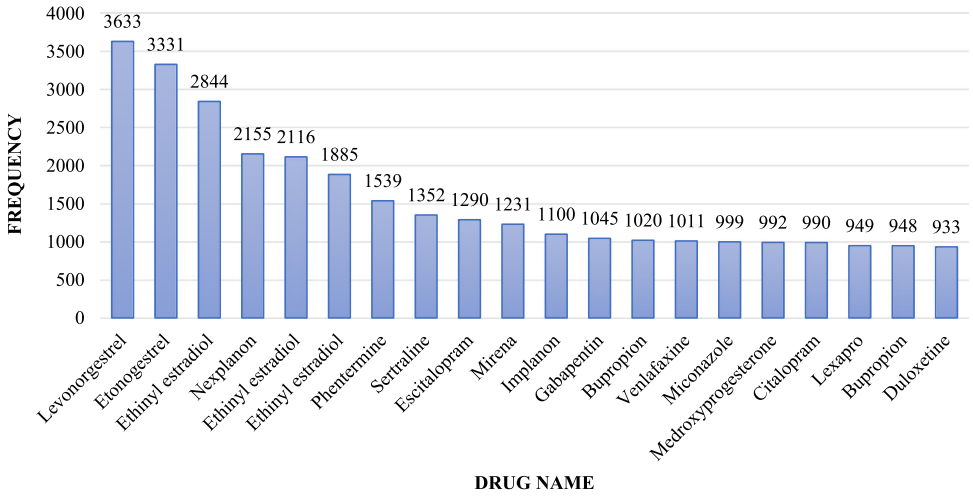

Fig. 9. Top 20 drug names in the dataset before normalization and preprocessing.

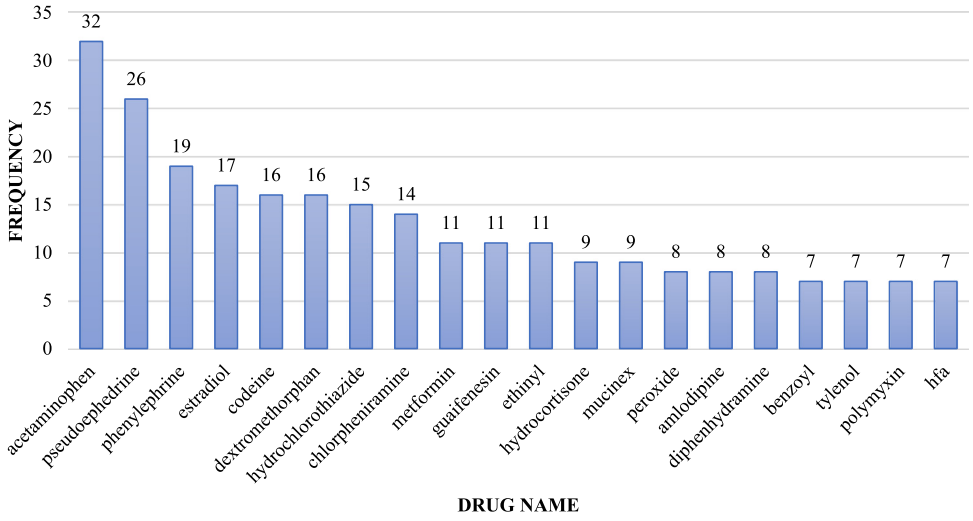

Fig. 10. Names of the top 20 drugs from the dataset after normalization and preprocessing.

classification models when try to classify hard samples in the dataset.

4. Experiments and results

This section presents the conducted experiments using our proposed fusion models applied to drug review data. This study introduces *two* fusion models by combining deep learning (GRU, CNN and 3CRNN) and traditional algorithms (NB, DT, RF and KNN) called 3-way fusion of one deep model with a traditional model (3W1DT) and 3-way fusion of *three* deep models with a traditional model (3W3DT) models. The obtained results using proposed fusion models and other applied methods are presented in the following subsections.

4.1. Data

In this study, we used a drug review data introduced by Gräßer et al. [24]. The dataset has 215063 instances in *three* main categories: positive, negative and natural. Sentiment classification (SA) has been widely investigated by using a variety of machine learning and deep learning methods in the literature [46–48]. Due to this importance, in this study, we attempted to show the power and effectiveness of machine learning and deep learning methods in SA of drug reviews.

Table 7
Comparison of the results obtained using the second fusion model for third experiment.

Method	Classes	Precision	Recall	F1-score
3W3DT-NB	Positive	0.8380	0.8582	0.8480
	Natural	0.9504	0.3808	0.5438
	Negative	0.8968	0.9618	0.9281
	weighted avg	**0.8868**	**0.8836**	**0.8735**
3W3DT-DT	Positive	0.8613	0.7108	0.7788
	Natural	0.6752	0.2881	0.4038
	Negative	0.8447	0.9670	0.9017
	weighted avg	0.8337	0.8417	0.8261
3W3DT-RF	Positive	0.8629	0.7222	0.7863
	Natural	0.6806	0.3009	0.4173
	Negative	0.8491	0.9663	0.9039
	weighted avg	0.8374	0.8453	0.8307
3W3DT-KNN	Positive	0.8635	0.6714	0.7554
	Natural	0.5813	0.2116	0.3103
	Negative	0.8244	0.9656	0.8894
	weighted avg	0.8124	0.8241	0.8038

4.2. Pre-processing

We used thresholds as defined in Table 4 to extract three-level polarity labels for the overall patient satisfaction ratings.

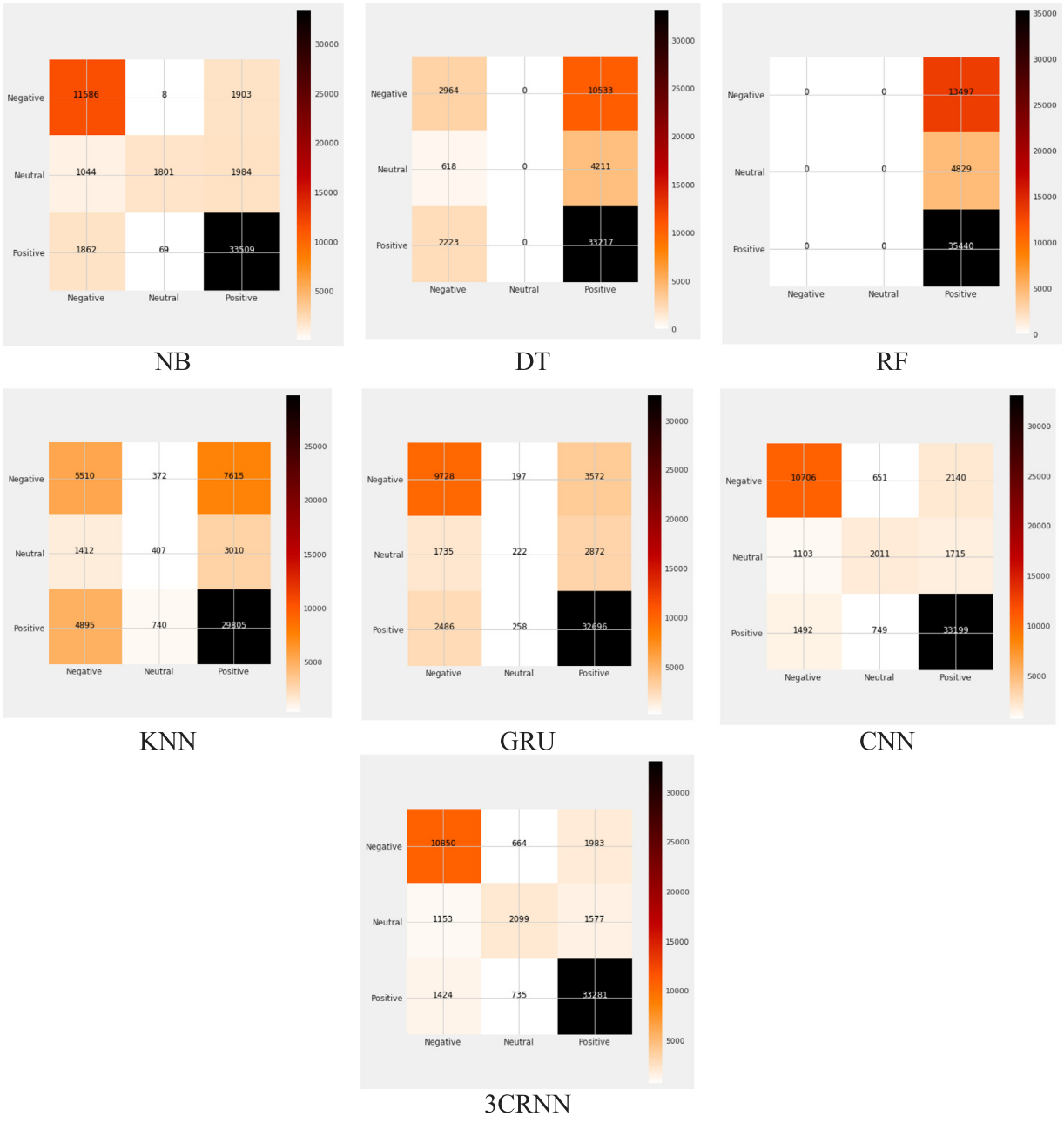

Fig. 11. Confusion matrices obtained using seven algorithms for the first experiment.

To remove outlier-length sentences from the dataset, we plotted the length of sentences against the number of sentences in Fig. 7(a). The average sentence length is 85.72 and the maximum is 1941. Moreover, the distribution of reviews by their length is shown in Fig. 8.

According to Fig. 8, we removed reviews of lengths greater than 180 words. The distribution of reviews based on their length in terms of words is shown in Fig. 8. The average and maximum sentence length are 85.25 and 180, respectively.

After performing the length normalization, we applied the following steps on the reviews:

- Tokenizing sentences to a list of separate words.
- Creating a stop-words dictionary.
- Adding the custom stop-words to predefined list of English stop-words.
- Creating a train–test split.

A large part of the corpus appeared to be drug names themselves at the preliminary examination. Instead of evaluating the individual names of medications and using them to construct a potentially biased weighting of names, we removed them by constructing a personalized product name corpus. Top 20 drug names obtained in the original dataset are shown in Figs. 9 and 10. This corpus is made by intersecting the Google 20k most common English words and normalized drug names in the dataset.

4.3. Preliminary results

In the first round of the experiments, four well-known classical algorithms, Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN), and three deep learning-based methods (GRU, CNN, and 3CRNN) are applied on the dataset. The confusion matrix of each method is shown in Fig. 11.

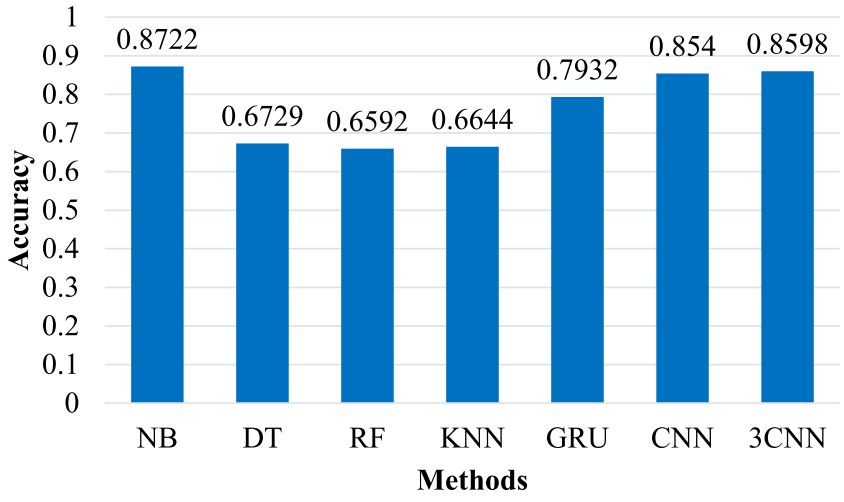

Fig. 12. Comparison of accuracies obtained using seven different algorithms for the first experiment.

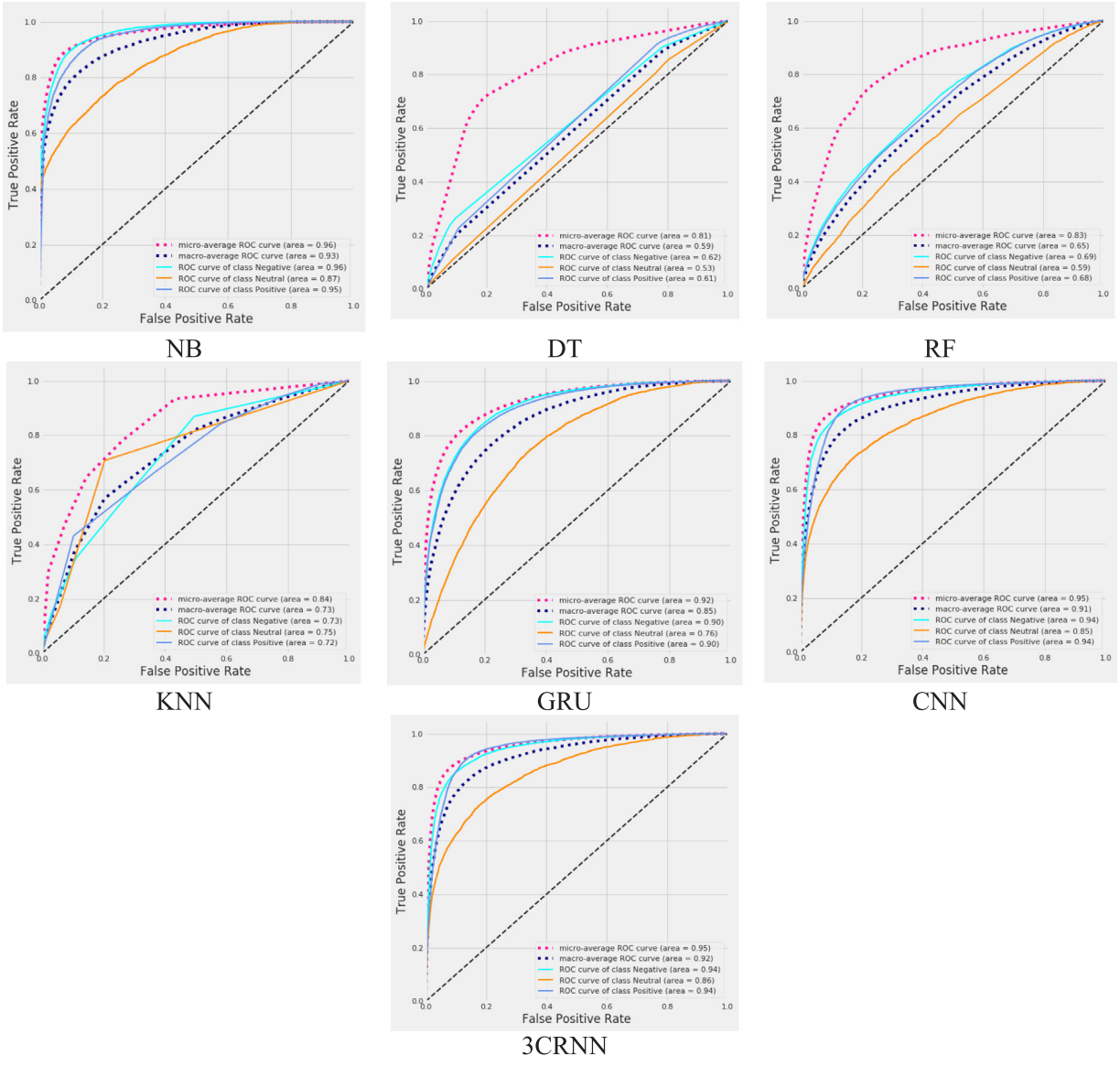

Fig. 13. ROCs obtained using seven algorithms for the first experiment.

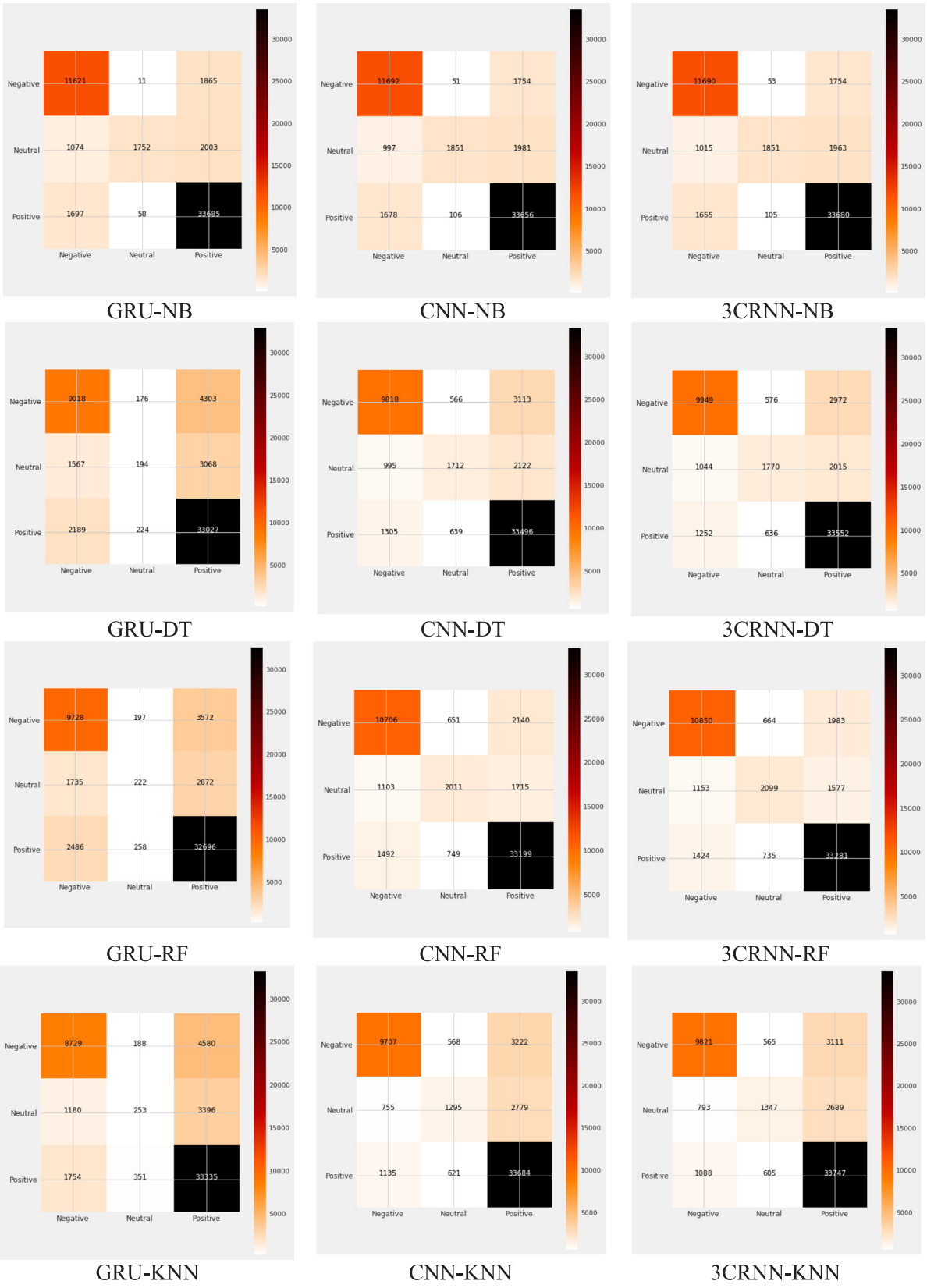

Fig. 14. Confusion matrices of the results obtained using the first fusion model for the second experiment.

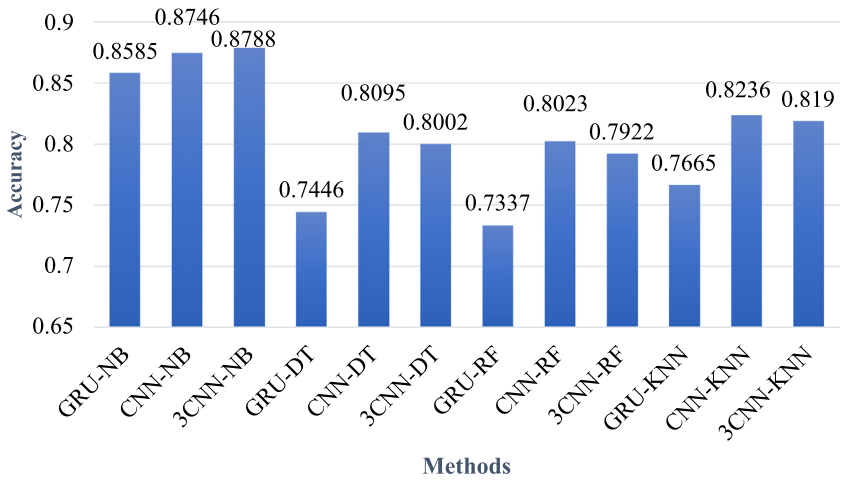

Fig. 15. Comparison of accuracies obtained for the second experiment using the proposed 3W1DT model.

Table 8
Comparison of the results obtained using the state-of-the-art deep learning methods.

Method	Classes	Precision	Recall	F1-score
AC-BiLSTM	Positive	0.7349	0.7685	0.7513
	Natural	0.4077	0.1510	0.2203
	Negative	0.8674	0.9267	0.8961
	weighted avg	0.7928	0.8173	0.7990
IWV	Positive	0.6967	0.7911	0.7409
	Natural	0.3808	0.1323	0.1964
	Negative	0.8767	0.9094	0.8928
	weighted avg	0.7870	0.8099	0.7921
CRNN	Positive	0.7514	0.8017	0.7758
	Natural	0.4640	0.3841	0.4203
	Negative	0.9057	0.9038	0.9047
	weighted avg	**0.8273**	**0.8315**	**0.8289**
ARC	Positive	0.7275	0.7164	0.7219
	Natural	0.3479	0.0839	0.1352
	Negative	0.8395	0.9312	0.8830
	weighted avg	0.7672	0.8012	0.7754

In Table 5, the algorithms are compared using four performance measures: precision, recall, F1-score and weighted average [52–55].

Moreover, to compare the algorithms in more details, the obtained accuracies and ROC measures [53] are presented in Figs. 12 and 13, respectively.

Table 1, Figs. 11 and 12 clearly illustrates that, NB performed very well as compared to other algorithms followed by 3CRNN and CNN. The achieved precision, recall, F1-score and accuracy of NB in the first experiment are 0.8775, 0.8722, 0.8626 and 0.8722, respectively. Moreover, it can be seen from Fig. 13 that, NB has yielded the best performance in terms of ROC metric as compared to other methods. As we discussed earlier, this study proposed two new fusion-based models by combining traditional machine learning algorithms with deep learning methods. The results obtained for each fusion model is presented in the following sub-sections.

4.4. Main results

In this section, the obtained results using two proposed fusion models are discussed. In the first step, 3-way fusion of one deep model with a traditional model (3W1DT) model is discussed followed by the second fusion model named 3-way fusion of three deep models with a traditional model (3W3DT).

4.4.1. 3-way fusion of one deep model with a traditional model (3W1DT)

In this step, to improve the performance of our applied methods, the first fusion model is proposed by combining traditional algorithms and deep learning methods. In the first fusion model (3W1DT) each classical algorithm is combined with all deep learning methods, separately. For example, NB is used with GRU, CNN, and 3CRNN termed as CNN-NB, GRU-NB and 3CRNN-NB. The obtained outcomes are presented in Fig. 14, Table 6, Figs. 15 and 16.

As mentioned earlier, the first proposed fusion model is used to improve the performance of various deep learning methods by using the classical algorithms. However, NB, DT, RF, and KNN are tested on the samples that deep learning methods could not classify them into one class with high confidence. According to the obtained results (Table 2, Figs. 15, and 16), it can be seen that the combination of NB with deep learning methods outperformed the other three traditional algorithms. More specifically, among applied deep learning methods, we found that combining 3CRNN and NB could achieve significantly outstanding performance as compared to the other methods. However, the improvements are not very significant. In this study, we tried to reduce the uncertainty in prediction level of deep learning and machine learning methods. In other words, in addition to the improvements, we had to deal with uncertainty issue using three-way

Table 9
Comparison of the performance of our best performing model (3W3DT-NB method) with other state-of-art techniques using the same database (drug review data).

Study	Year	Method	Measures			
			Precision	Recall	F1-score	Accuracy (%)
Gräßer et al. [24]	2018	logistic regression	–	–	–	69.88
Colón–Ruizet al. [49]	2019	CNN (w2ventrenable)	0.6672	0.6672	0.6672	–
Jain et al. [50]	2019	Deep Neural Network(DNN)	0.8400	0.8300	–	83.80
Chen et al. [51]	2019	Fuzzy-rough featureSelection + bag ofwords (BoW) + Ripper	–	–	–	65.06
This study	**2020**	**3W3DT-NB**	**0.8868**	**0.8836**	**0.8735**	**88.36**

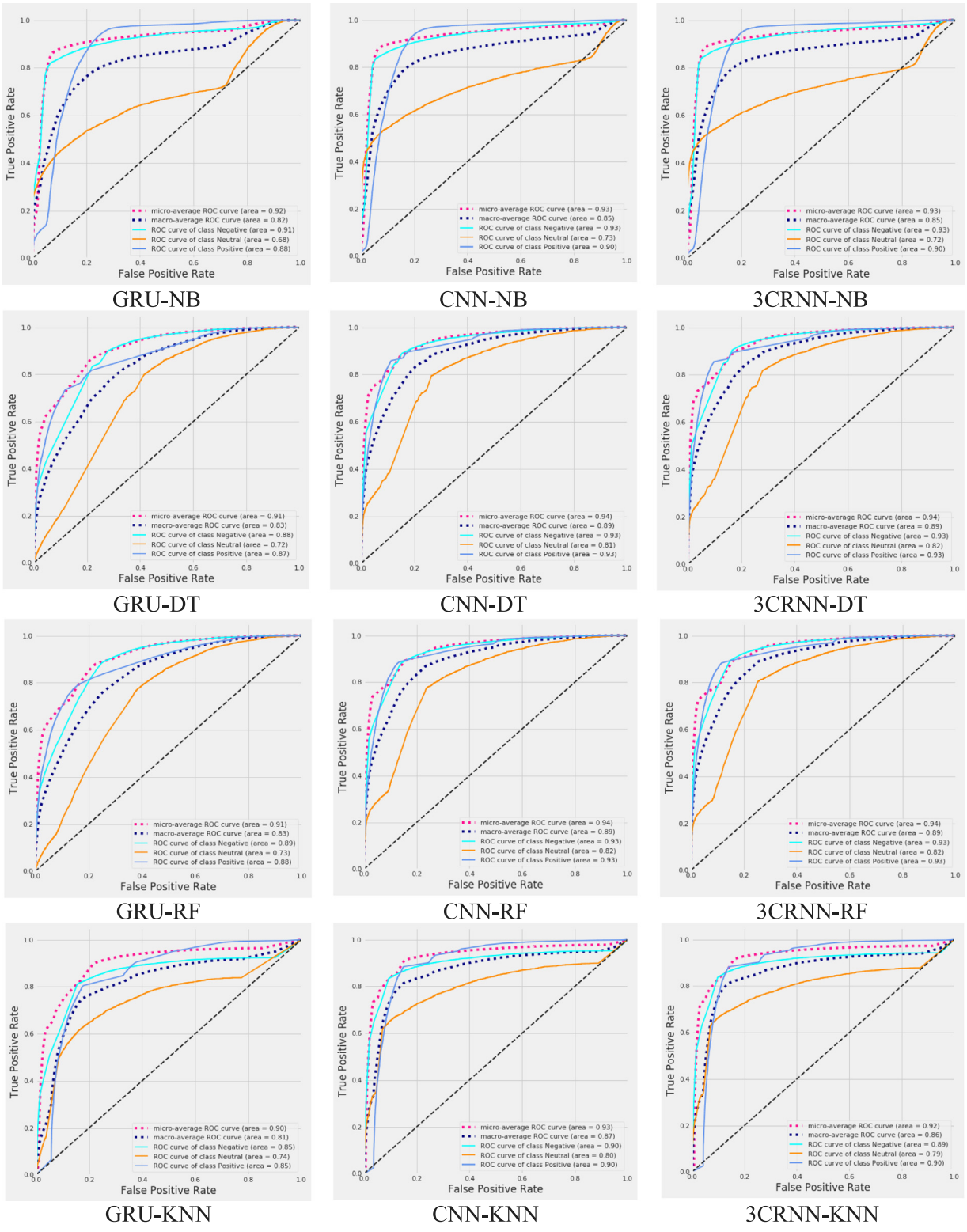

Fig. 16. ROCs obtained using seven algorithms for the second experiment with the proposed 3W1DT model.

approach. This means that, we identified the samples for which models are unable to classify correctly and fed them to another classifier. In other words, external classifier contributed to classify correctly the misclassified samples. For further improvement of drug review classification, we proposed the second fusion model called 3-way fusion of three deep models with a traditional model (3W3DT).

4.4.2. 3-way fusion of three deep models with a traditional model (3W3DT),

In this section, the second fusion model, called **3-way fusion of three deep models with a traditional model (3W3DT)**, is employed to improve the performance of deep learning methods. In this model, *three* deep learning algorithms (GRU, CNN and 3CRNN) together with traditional algorithms (NB, DT, RF

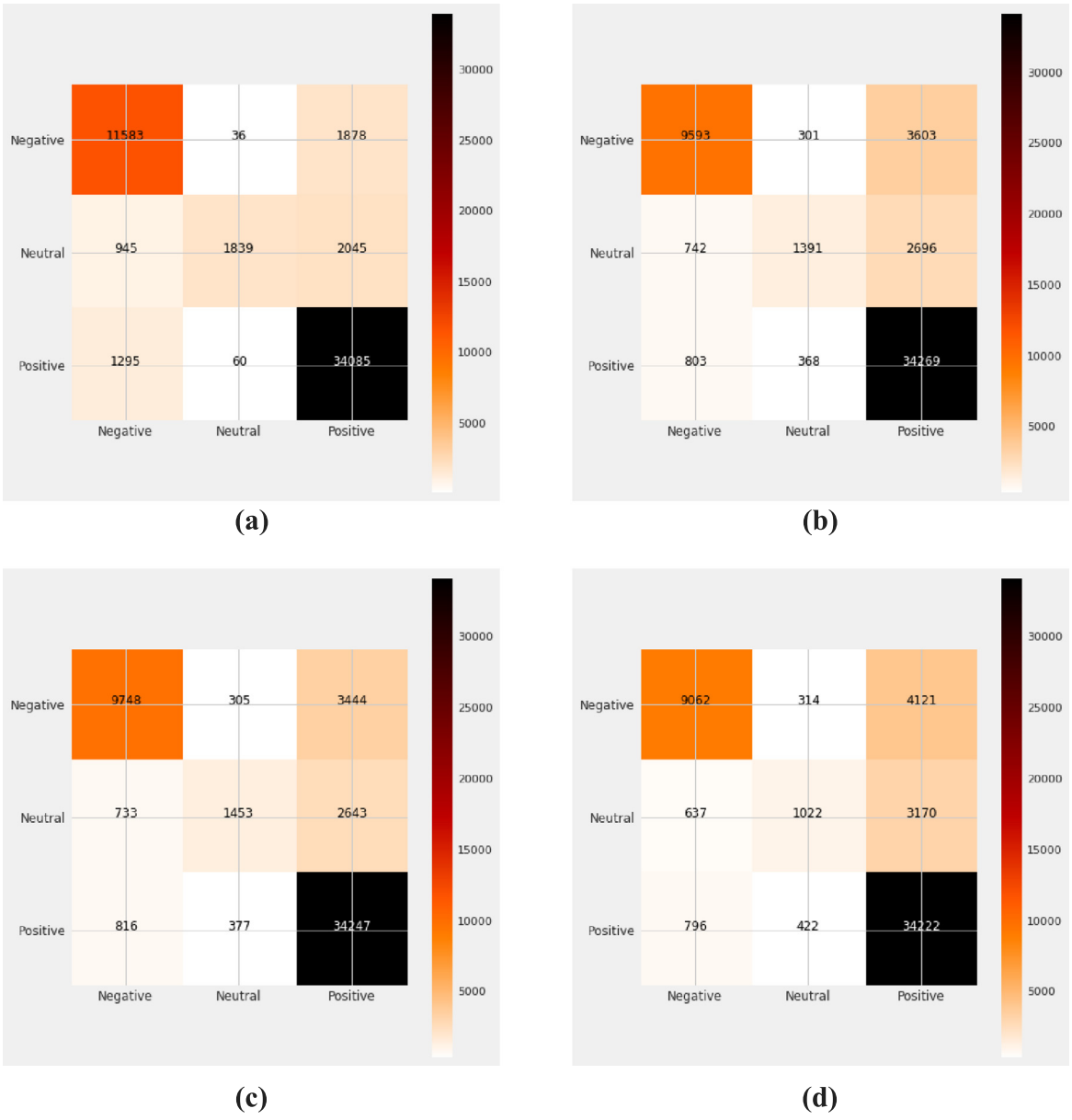

Fig. 17. Confusion matrices of the results obtained using the second fusion model for third experiment employing (a) 3W3DT-NB, (b) 3W3DT-DT, (c) 3W3DT-RF and (d) 3W3DT-KNN methods.

and KNN) are used one by one. The new hybrid methods called **3W3DT**-NB, **3W3DT**-DT, **3W3DT**-RF and **3W3DT**-KNN are employed. The obtained results using our second fusion model are presented in Fig. 17, Table 7, Figs. 18, and 19.

It can be noted from Table 7, Figs. 18, and 19 that, **3W3DT** model significantly improved the results. In the previous steps, the combination of NB and deep learning methods achieved better outcomes as compared to the other used methods to SA in drug reviews. Interestingly, we found that even though RF in the second and third experiments did not yield good performances, it obtained slightly better than DT. In contrast, KNN in the second experiment (3W1DT) showed better performance as compared to the first experiment. The performance improvement of KNN with 3W1DT ranked the second best method while with **3W3DT** ranked the fourth best method. These outcomes demonstrate that the combination of NB and deep learning performed well for SA with drug data.

4.5. Comparison with state-of-the-art deep learning methods

In order to show the effectiveness of the proposed fusion models, we implemented four state-of-the-art deep learning methods. We have used drug review data on AC-BiLSTM [56], Improved Word Vectors (IWV) [57], CRNN [58], and attention neural network (ARC) [59] models. The obtained results are presented in Fig. 20, Table 8, Figs. 21, and 22.

It can be noted from the results that CRNN showed better performance as compared to other three applied methods. CRNN achieved the precision, recall, F1-score, and accuracy of 0.8273, 0.8315, 0.8289, and 0.8315, respectively. In contrast, ARC obtained the weakest performance as compared to the other state-of-the-art deep learning methods with precision, recall, F1-score, and accuracy of 0.7672, 0.8012, 0.7754, and 0.8012, respectively.

We have also compared the performance of our second fusion model with other studies which used the same dataset (see Table 9).

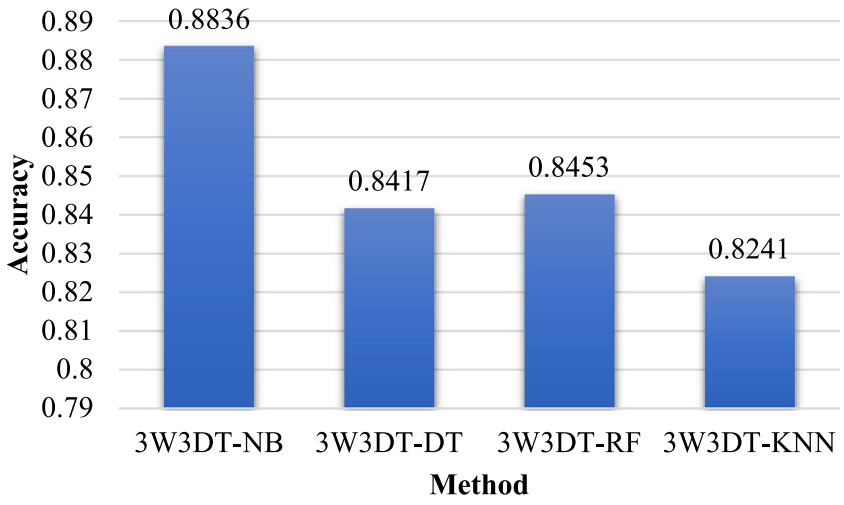

Fig. 18. Comparison of accuracies obtained with the third experiment using the proposed 3W3DT model.

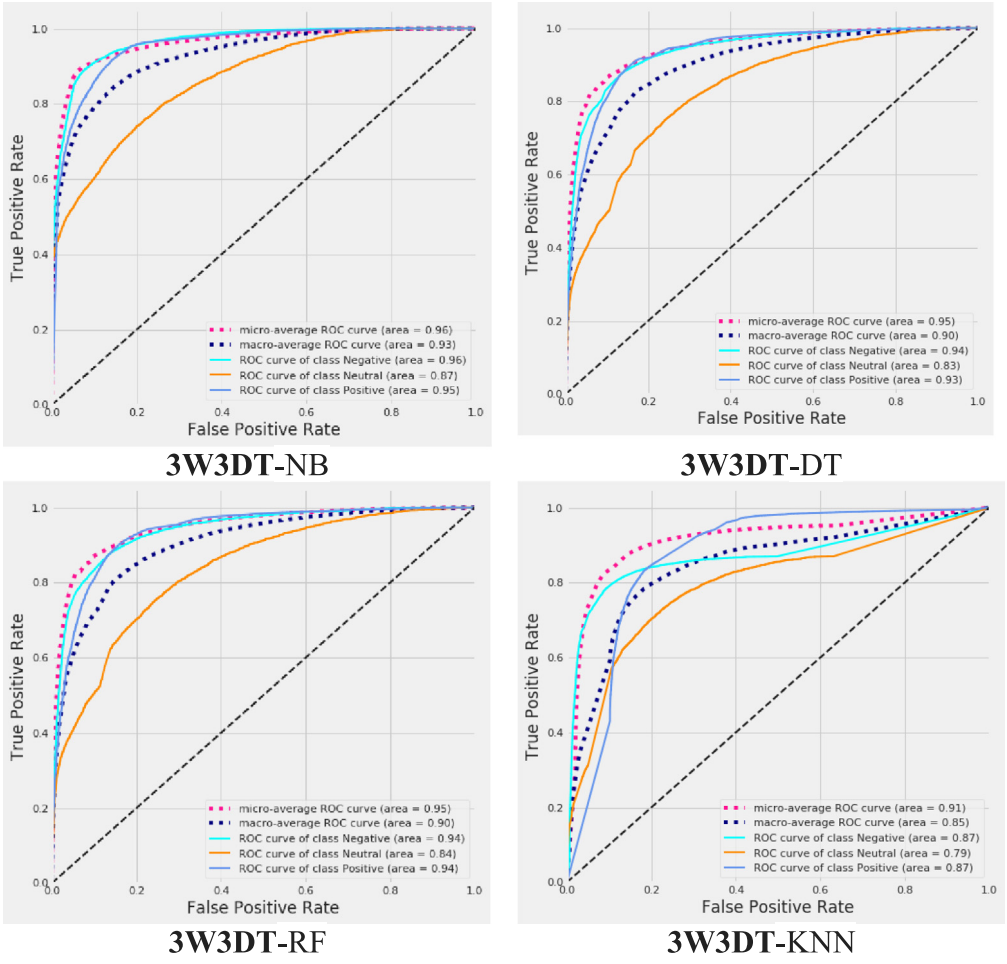

Fig. 19. ROCs for seven algorithms employed for the third experiment using the proposed 3W3DT model.

It can be noted from Table 5 that, our second proposed fusion model (3W3DT-NB) performed better than other existing methods in the literature.

The main advantages of our fusion models are listed below.

1. Takes into consideration the misclassified samples.
2. Obtained outstanding performance.
3. Considered uncertainty quantification (UQ) in SA.

In the future, we plan to improve the performance of our proposed fusion models using various feature selections methods [60] based on metaheuristic algorithms [61]. To do so, our first plan is to apply recently introduced metaheuristic and evolutionary-based algorithms [62–64]. Moreover, the strength

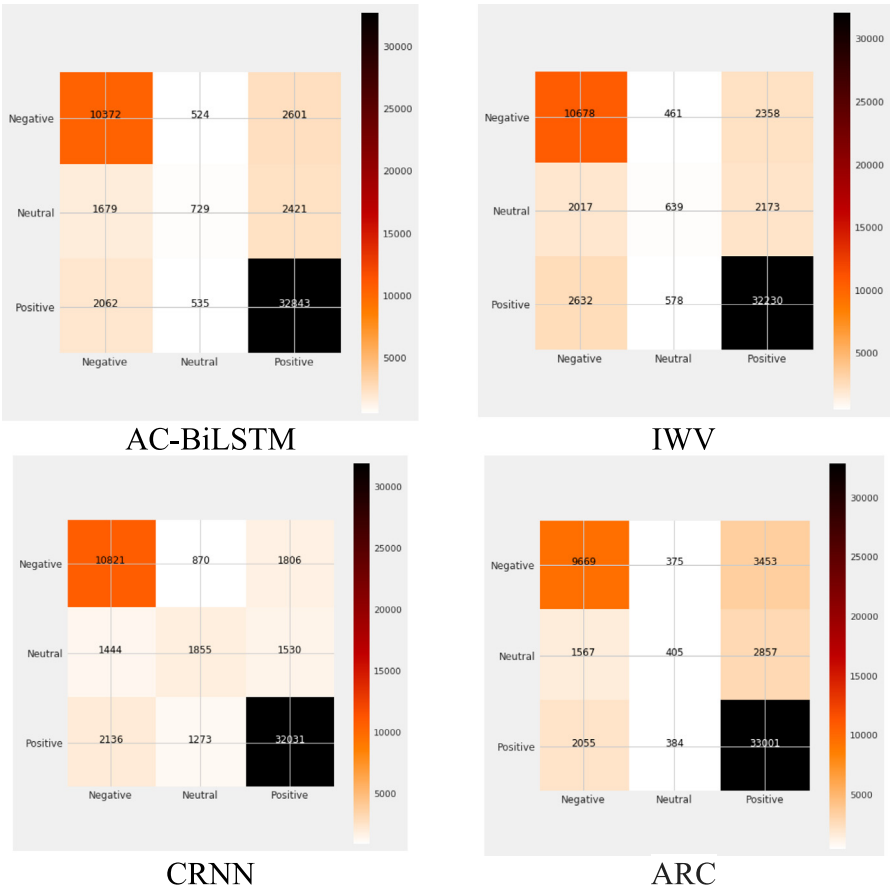

Fig. 20. Confusion matrices of the results obtained using the applied state-of-the-art deep learning methods.

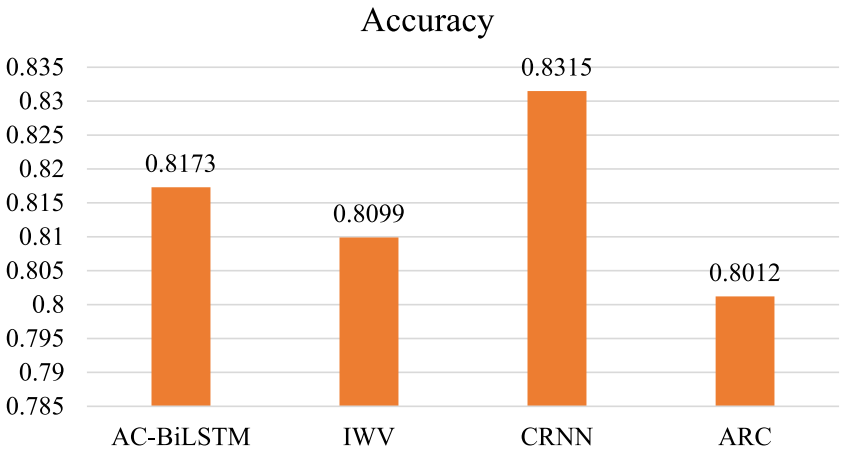

Fig. 21. Comparison of accuracies obtained in the third experiment using the state-of-the-art deep learning methods.

of Q-learning in ensemble learning is studied in [65]. Hence, we intend to apply Q-learning algorithm (i.e., a Q-learning based dynamic model selection (QDMS) [66], Q-learning based on multi-agent classifier system (QMACS) [65]). Moreover, we plan to introduce weights to each model instead of using them with the same weight. To do so, we aim to apply estimation algorithms (EAs) to find the most appropriate weight in each method [53,67]. In addition, we aim to use other recent proposed deep learning based methods and fusion techniques such as hierarchica deep genetic networks [68], transfer learning-based deep models [69], and other feature extraction methods using both textual and image medical data [70–73].

5. Conclusion

In this study, we proposed *two* fusion sentiment classification models called 3-way fusion of one deep model with a traditional model (3W1DT) and 3-way fusion of *three* deep models with a traditional model (3W3DT). In 3W1DT a deep learning method is utilized as the base classifier (BC) which is applied to classify the test samples in the high-confidence region. However, in 3W3DT *three* deep and one traditional models are trained on the entire training data and each classifies the test sample individually. Moreover, we have compared our results with *four* state-of-the-art deep learning methods: AC-BiLSTM, improved word vectors

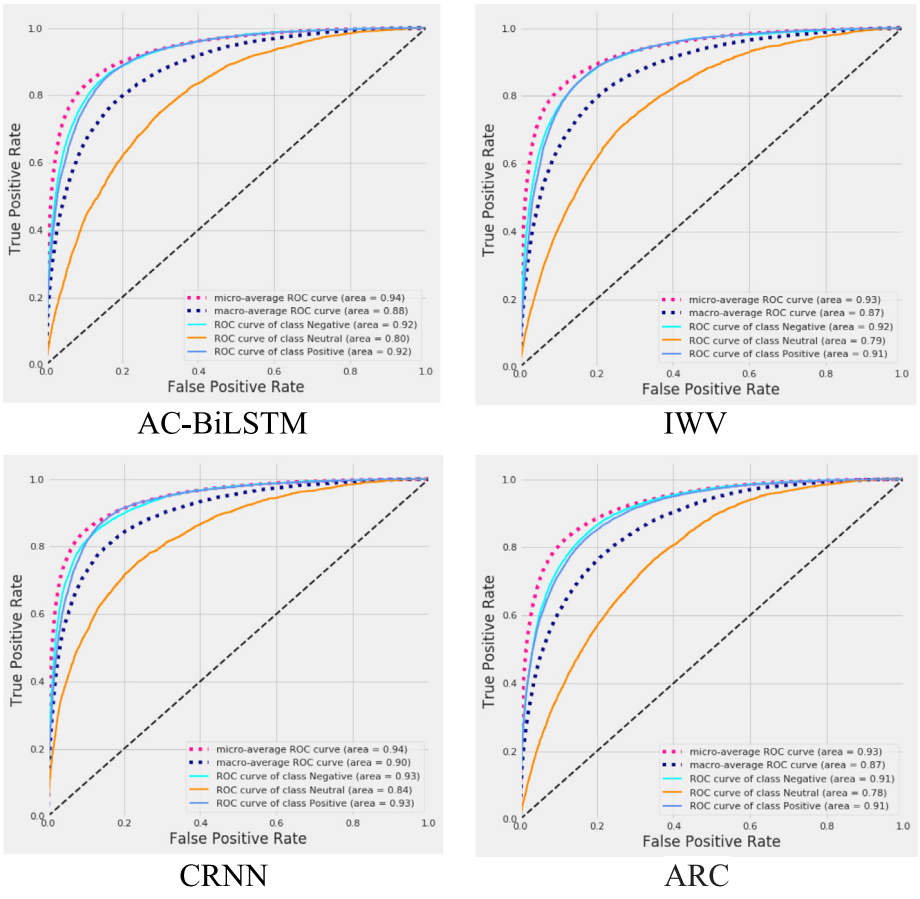

Fig. 22. ROCs obtained using seven algorithms for third experiment with state-of-the-art deep learning methods.

(IWV), CRNN, and attention neural network (ARC) by implementing these models. The result of the second fusion model (3W3DT) outperformed our first proposed fusion model (3W1DT) and all *four* state-of-the-art deep learning methods (AC-BiLSTM, IWV, CRNN and ARC). The proposed models may be applied to other similar classification problems such as emotion recognition and rating prediction in the domain of SA. However, in order to apply the proposed method to such problems, the confidence calculation mechanism should be adopted for the specific problem. Moreover, the boundaries in the sample space must be considered carefully to divide the sample space into appropriate number of classes. In future, we intend to use this proposed model for other biomedical applications such as managing healthcare records, detection of cardiac and neural diseases.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Mohammad Ehsan Basiri: Conceptualization, Methodology, Software, Writing - original draft. **Moloud Abdar:** Data curation, Conceptualization, Writing - original draft, Writing - review & editing. **Mehmet Akif Cifci:** Writing - original draft, Investigation. **Shahla Nemati:** Conceptualization, Methodology. **U. Rajendra Acharya:** Supervision, Validation, Writing - review & editing.

Acknowledgments

This work was financially supported by the research deputy of Shahrekord University (grant 98GRD30M1874).

References

[1] S.M. Jiménez-Zafra, M.T. Martín-Valdivia, M.D. Molina-González, L.A. Ureña-López, How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for the medical domain, Artif. Intell. Med. 93 (2019) 50–57.
[2] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2) (2016) 102–107.
[3] B. Ru, D. Li, Y. Hu, L. Yao, Serendipity—A machine-learning application for mining serendipitous drug usage from social media, IEEE Trans. NanoBioscience 18 (3) (2019) 324–334.
[4] M.E. Basiri, A. Kabiri, Translation is not enough: comparing lexicon-based methods for sentiment analysis in persian, in: 2017 International Symposium on Computer Science and Software Engineering Conference, CSSE, IEEE, 2017, pp. 36–41.
[5] R.M. Duwairi, N.A. Ahmed, S.Y. Al-Rifai, Detecting sentiment embedded in arabic social media–a lexicon-based approach, J. Intell. Fuzzy Systems 29 (1) (2015) 107–117.
[6] B. Zhang, D. Xu, H. Zhang, M. Li, STCS lexicon: Spectral-clustering-based topic-specific Chinese sentiment lexicon construction for social networks, IEEE Trans. Comput. Soc. Syst. 6 (6) (2019) 1180–1189.
[7] F.M. Plaza-del Arco, M.T. Martín-Valdivia, L.A. Ureña-López, R. Mitkov, Improved emotion recognition in spanish social media through incorporation of lexical knowledge, Future Gener. Comput. Syst. (2019).
[8] M.Z. Asghar, A. Khan, S. Ahmad, M. Qasim, I.A. Khan, Lexicon-enhanced sentiment analysis framework using rule-based classification scheme, PLoS One 2 (2) (2017).
[9] S. Vashishtha, S. Susan, Fuzzy rule based unsupervised sentiment analysis from social media posts, Expert Syst. Appl. 138 (2019) 112834.
[10] M.E. Basiri, M. Abdar, A. Kabiri, S. Nemati, X. Zhou, F. Allahbakhshi, N.Y. Yen, Improving sentiment polarity detection through target identification, IEEE Trans. Comput. Soc. Syst. (2019).

[11] M.E. Basiri, A. Kabiri, Words are important: improving sentiment analysis in the persian language by lexicon refining, ACM Trans. Asian Low-Resour. Lang. Inf. Process. 17 (4) (2018) 1–8.
[12] M.E. Basiri, A. Kabiri, HOMPer: A new hybrid system for opinion mining in the Persian language, J. Inf. Sci. (2019).
[13] Z. Teng, D.T. Vo, Y. Zhang, Context-sensitive lexicon features for neural sentiment analysis, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 1629–1638.
[14] K.F. Hew, X. Hu, C. Qiao, Y. Tang, What predicts student satisfaction with MOOCs: a gradient boosting trees supervised machine learning and sentiment analysis approach, Comput. Educ. 145 (2020) 103724.
[15] B.S. Rintyarna, R. Sarno, C. Fatichah, Evaluating the performance of sentence level features and domain sensitive features of product reviews on supervised sentiment analysis tasks, J. Big Data 6 (1) (2019) 84.
[16] C. Wu, F. Wu, S. Wu, Z. Yuan, J. Liu, Y. Huang, Semi-supervised dimensional sentiment analysis with variational autoencoder, Knowl.-Based Syst. 165 (2019) 30–39.
[17] Y. Han, Y. Liu, Z. Jin, Sentiment analysis via semi-supervised learning: a model based on dynamic threshold and multi-classifiers, Neural Comput. Appl. (2019) 1–3.
[18] S. Poria, E. Cambria, N. Howard, G.B. Huang, Hussain A., Fusing audio visual and textual clues for sentiment analysis from multimodal content, Neurocomputing 174 (2016) 50–59.
[19] Z. Zhao, H. Zhu, Z. Xue, Z. Liu, J. Tian, M.C. Chua, M. Liu, An image-text consistency driven multimodal sentiment analysis approach for social media, Inf. Process. Manage. 56 (6) (2019) 102097.
[20] N. Majumder, D. Hazarika, A. Gelbukh, E. Cambria, S. Poria, Multimodal sentiment analysis using hierarchical fusion with context modeling, Knowl.-Based Syst. 161 (2018) 124–133.
[21] S. Nemati, R. Rohani, M.E. Basiri, M. Abdar, N.Y. Yen, V. Makarenkov, A hybrid latent space data fusion method for multimodal emotion recognition, IEEE Access 7 (2019) 172948–172964.
[22] M. Ebrahimi, A.H. Yazdavar, N. Salim, S. Eltyeb, Recognition of side effects as implicit-opinion words in drug reviews, Online Inf. Rev. (2016).
[23] S.W. Chew, C.S. Khoo, Comparison of drug information on consumer drug review sites versus authoritative health information websites, J. Assoc. Inf. Sci. Technol. 67 (2) (2016) 333–349.
[24] F. Gräßer, S. Kallumadi, H. Malberg, S. Zaunseder, Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning, in: Proceedings of the 2018 International Conference on Digital Health, ACM, 2018, pp. 112–125.
[25] S.M. Jiménez-Zafra, M.T. Martín-Valdivia, M.D. Molina-González, L.A. Ureña-López, How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain, Artif. Intell. Med. 93 (2019) 50–57.
[26] S. Liu, I. Lee, Extracting features with medical sentiment lexicon and position encoding for drug reviews, Health Inf. Sci. Syst. 7 (1) (2019) 11.
[27] M. Zhang, G. Geng, Adverse drug event detection using a weakly supervised convolutional neural network and recurrent neural network model, Information 10 (9) (2019) 276.
[28] Y. Han, M. Liu, W. Jing, Aspect-level drug reviews sentiment analysis based on double BiGRU and knowledge transfer, IEEE Access (2020).
[29] H.H. Do, P.W.C. Prasad, A. Maag, A. Alsadoon, Deep learning for aspect-based sentiment analysis: a comparative review, Expert Syst. Appl. 118 (2019) 272–299.
[30] T. Young, D. Hazarika, S. Poria, E. Cambria, Recent trends in deep learning based natural language processing, IEEE Comput. Intell. Mag. 13 (3) (2018) 55–75.
[31] L. Zhang, S. Wang, B. Liu, Deep learning for sentiment analysis: A survey, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 8 (4) (2018) e1253.
[32] F. Liu, J. Zheng, L. Zheng, C. Chen, Combining attention-based bidirectional gated recurrent neural network and two-dimensional convolutional neural network for document-level sentiment classification, Neurocomputing 371 (2020) 39–50.
[33] A. Kumar, K. Srinivasan, W.H. Cheng, A.Y. Zomaya, Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data, Inf. Process. Manage. 57 (1) (2020) 102141.
[34] B. Fan, W. Fan, C. Smith, Adverse drug event detection and extraction from open data: A deep learning approach, Inf. Process. Manage. 57 (1) (2020) 102131.
[35] H.J. Park, M. Song, K.S. Shin, Deep learning models and datasets for aspect term sentiment classification: Implementing holistic recurrent attention on target-dependent memories, Knowl.-Based Syst. 187 (2020) 104825.
[36] S. Minaee, E. Azimi, A. Abdolrashidi, Deep-sentiment: Sentiment analysis using ensemble of CNN and bi-LSTM models, 2019, arXiv preprint arXiv:1904.04206.
[37] Y. Ma, H. Peng, E. Cambria, Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
[38] J. Na, W.Y.M. Kyaing, Sentiment analysis of user-generated content on drug review websites, J. Inf. Sci. Theory Pract. 3 (1) (2015) 6–23.
[39] K. Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio, On the properties of neural machine translation: Encoder-decoder approaches, 2014, arXiv preprint arXiv:1409.1259.
[40] Y. Yao, Three-way decisions with probabilistic rough sets, Inform. Sci. 180 (3) (2010) 341–353.
[41] Y. Zhang, Z. Zhang, D. Miao, J. Wang, Three-way enhanced convolutional neural networks for sentence-level sentiment classification, Inform. Sci. 477 (2019) 55–64.
[42] I. Chaturvedi, E. Cambria, R.E. Welsch, F. Herrera, Distinguishing between facts and opinions for sentiment analysis: Survey and challenges, Inf. Fusion 44 (2018) 65–77.
[43] Z. Li, Y. Fan, B. Jiang, T. Lei, W. Liu, A survey on sentiment analysis and opinion mining for social multimedia, Multimedia Tools Appl. 78 (6) (2019) 6939–6967.
[44] J.C. Na, W.Y. Kyaing, C.S. Khoo, S. Foo, Y.K. Chang, Y.L. Theng, Sentiment classification of drug reviews using a rule-based linguistic approach, in: International Conference on Asian Digital Libraries, Springer, Berlin, Heidelberg, 2012, pp. 189–198.
[45] W. Rawat, Z. Wang, Deep convolutional neural networks for image classification: A comprehensive review, Neural Comput. 9 (9) (2017) 2352–2449.
[46] A. Tripathy, A. Agrawal, S.K. Rath, Classification of sentiment reviews using n-gram machine learning approach, Expert Syst. Appl. 57 (2016) 117–126.
[47] O. Araque, I. Corcuera-Platas, J.F. Sánchez-Rada, C.A. Iglesias, Enhancing deep learning sentiment analysis with ensemble techniques in social applications, Expert Syst. Appl. 77 (2017) 236–246.
[48] S. Seo, C. Kim, H. Kim, K. Mo, P. Kang, Comparative study of deep learning-based sentiment classification, IEEE Access 8 (2020) 6861–6875.
[49] C. Colón-Ruiz, I. Segura-Bedmar, P. Martínez, Análisis de Sentimiento en el dominio salud: analizando comentarios sobre fármacos, Proces. Leng. Nat. 63 (2019) 15–22.
[50] N. Jain, A. Kumar, S. Singh, C. Singh, S. Tripathi, Deceptive reviews detection using deep learning technique, in: International Conference on Applications of Natural Language to Information Systems, Springer, Cham, 2019, pp. 79–91.
[51] T. Chen, P. Su, C. Shang, R. Hill, H. Zhang, Q. Shen, Sentiment classification of drug reviews using fuzzy-rough feature selection, in: 2019 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE, IEEE, 2019, pp. 1–6.
[52] M. Abdar, V. Makarenkov, CWV-BANN-SVM Ensemble learning classifier for an accurate diagnosis of breast cancer, Measurement 146 (2019) 557–570.
[53] A. Koohestani, M. Abdar, A. Khosravi, S. Nahavandi, M. Koohestani, Integration of ensemble and evolutionary machine learning algorithms for monitoring diver behavior using physiological signals, IEEE access 7 (2019) 98971–98992.
[54] M. Abdar, U.R. Acharya, N. Sarrafzadegan, V. Makarenkov, NE-nu-SVC: A new nested ensemble clinical decision support system for effective diagnosis of coronary artery disease, IEEE Access 7 (2019) 167605–167620.
[55] M. Abdar, V.N. Wijayaningrum, S. Hussain, R. Alizadehsani, P. Plawiak, U.R. Acharya, V. Makarenkov, IAPSO-AIRS: A novel improved machine learning-based system for wart disease treatment, J. Med. Syst. 43 (7) (2019) 220.
[56] G. Liu, J. Guo, Bidirectional LSTM with attention mechanism and convolutional layer for text classification, Neurocomputing 337 (2019) 325–338.
[57] S.M. Rezaeinia, R. Rahmani, A. Ghodsi, H. Veisi, Sentiment analysis based on improved pre-trained word embeddings, Expert Syst. Appl. 117 (2019) 139–147.
[58] X. Wang, W. Jiang, Z. Luo, Combination of convolutional and recurrent neural network for sentiment analysis of short texts, in: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016, pp. 2428–2437.
[59] S. Wen, J. Li, Recurrent convolutional neural network with attention for Twitter and yelp sentiment classification: ARC model for sentiment classification, in: Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence, ACM, 2018, p. 49.
[60] M. Asif, A. Ishtiaq, H. Ahmad, H. Aljuaid, J. Shah, Sentiment analysis of extremism in social media from textual information, Telemat. Inform. (2020) 101345.
[61] A. Jain, B.P. Nandi, C. Gupta, D.K. Tayal, Senti-NSetPSO: large-sized document-level sentiment analysis using neutrosophic set and particle swarm optimization, Soft Comput. 24 (1) (2020) 3–15.
[62] F. Pourpanah, Y. Shi, C.P. Lim, Q. Hao, C.J. Tan, Feature selection based on brain storm optimization for data classification, Appl. Soft Comput. 80 (2019) 761–775.
[63] F. Pourpanah, C.P. Lim, X. Wang, C.J. Tan, M. Seera, Y. Shi, A hybrid model of fuzzy min–max and brain storm optimization for feature selection and data classification, Neurocomputing 333 (2019) 440–451.

[64] R. Stoean, C. Stoean, Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection, Expert Syst. Appl. 40 (7) (2013) 2677–2686.
[65] F. Pourpanah, R. Wang, C.P. Lim, X. Wang, M. Seera, C.J. Tan, An improved fuzzy ARTMAP and Q-learning agent model for pattern classification, Neurocomputing 359 (2019) 139–152.
[66] C. Feng, M. Sun, J. Zhang, Reinforced deterministic and probabilistic load forecasting via Q-learning dynamic model selection, IEEE Trans. Smart Grid (2019).
[67] F. Pourpanah, C.P. Lim, J.M. Saleh, A hybrid model of fuzzy ARTMAP and genetic algorithm for data classification and rule extraction, Expert Syst. Appl. 49 (2016) 74–85.
[68] P. Pławiak, M. Abdar, J. Pławiak, V. Makarenkov, U.R. Acharya, DGHNL: A new deep genetic hierarchical network of learners for prediction of credit scoring, Inform. Sci. 516 (2020) 401–418.
[69] Y. Celik, M. Talo, O. Yildirim, M. Karabatak, U.R. Acharya, Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images, Pattern Recognit. Lett. (2020).
[70] O.S. Lih, V. Jahmunah, T.R. San, E.J. Ciaccio, T. Yamakawa, M. Tanabe, M. Kobayashi, O. Faust, U.R. Acharya, Comprehensive electrocardiographic diagnosis based on deep learning, Artif. Intell. Med. 103 (2020) 101789.
[71] V. Jahmunah, S.L. Oh, V. Rajinikanth, E.J. Ciaccio, K.H. Cheong, N. Arunkumar, U.R. Acharya, Automated detection of schizophrenia using nonlinear signal processing methods, Artif. Intell. Med. 100 (2019) 101698.
[72] U.R. Acharya, S.L. Fernandes, J.E. WeiKoh, E.J. Ciaccio, M.K. Fabell, U.J. Tanik, V. Rajinikanth, C.H. Yeong, Automated detection of Alzheimer's disease using brain MRI images–a study with various feature extraction techniques, J. Med. Syst. 43 (9) (2019) 302.
[73] A. Bhandary, G.A. Prabhu, V. Rajinikanth, K.P. Thanaraj, S.C. Satapathy, D.E. Robbins, C. Shasky, Y.D. Zhang, J.M. Tavares, N.S. Raja, Deep-learning framework to detect lung abnormality–A study with chest X-ray and lung CT scan images, Pattern Recognit. Lett. 129 (2020) 271–278.