

# Sentiment Analysis of Persian Movie Reviews Using Deep Learning

Kia Dashtipour <sup>1,\*</sup>, Mandar Gogate <sup>2</sup>, Ahsan Adeel <sup>3</sup>, Hadi Larijani <sup>4</sup>  and Amir Hussain <sup>2</sup>

<sup>1</sup> Department of Computing Science and Mathematics, University of Stirling, Stirling FK9 4LA, UK

<sup>2</sup> School of Computing, Edinburgh Napier University, Edinburgh EH11 4BN, UK; m.gogate@napier.ac.uk (M.G.); Amir.Hussain@napier.ac.uk (A.H.)

<sup>3</sup> School of Mathematics and Computer Science, University of Wolverhampton, Wolverhampton WV1 1LY, UK; aadeel200@gmail.com

<sup>4</sup> School of Computing, Engineering and Built Environment, Glasgow Caledonian University, Glasgow G4 0BA, UK; H.Larijani@gcu.ac.uk

\* Correspondence: k.dashtipour@napier.ac.uk

**Abstract:** Sentiment analysis aims to automatically classify the subject's sentiment (e.g., positive, negative, or neutral) towards a particular aspect such as a topic, product, movie, news, etc. Deep learning has recently emerged as a powerful machine learning technique to tackle the growing demand for accurate sentiment analysis. However, the majority of research efforts are devoted to English-language only, while information of great importance is also available in other languages. This paper presents a novel, context-aware, deep-learning-driven, Persian sentiment analysis approach. Specifically, the proposed deep-learning-driven automated feature-engineering approach classifies Persian movie reviews as having positive or negative sentiments. Two deep learning algorithms, convolutional neural networks (CNN) and long-short-term memory (LSTM), are applied and compared with our previously proposed manual-feature-engineering-driven, SVM-based approach. Simulation results demonstrate that LSTM obtained a better performance as compared to multilayer perceptron (MLP), autoencoder, support vector machine (SVM), logistic regression and CNN algorithms.



**Citation:** Dashtipour, K.; Gogate, M.; Adeel, A.; Larijani, H.; Hussain, A. Sentiment Analysis of Persian Movie Reviews Using Deep Learning. *Entropy* **2021**, *23*, 596. <https://doi.org/10.3390/e23050596>

Received: 4 March 2021

Accepted: 4 May 2021

Published: 12 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** sentiment analysis; deep learning; CNN; LSTM; classification

## 1. Introduction

Sentiment analysis is a method to automatically classify large amounts of text into positive or negative sentiments [1–4]. With the explosive growth of social media, organizations and companies have started to use big online data for their product enhancement and proactive decision making. In recent years, social media, forums, blogs, and other forms of online communication tools have radically affected everyday life, especially how people express their opinions. The extraction of useful information (for example, people's opinions about company brands) from the huge amount of unstructured data is important for most companies and organizations [5–9]. The application of sentiment analysis is not limited to product or movie reviews. Researchers have also applied it to areas such as news, politics, sport, etc. For example, in online political debates, sentiment analysis can be used to identify people's opinions about a certain candidate or political party [10–16]. However, sentiment analysis has been widely used for the English language using traditional and advanced machine learning techniques, and limited research has been conducted to develop models for the Persian language [17–20].

In the literature, Sentiment analysis has been performed at both the document and sentence level. The document-level sentiment analysis has been used to classify the sentiments expressed in the document (positive or negative), whereas, at the sentence level, the models have been used to identify the sentiments expressed only in the sentence under analysis [21–27]. For sentiment analysis, there are two widely used approaches: (1) Lexicon-based approach that uses lexicons (a dictionary of words and corresponding polarities) to

assign polarity; (2) Machine learning approach which requires a large labelled dataset with manual annotation [28–32]. Recently, deep-machine-learning-based automated feature engineering and classification has been shown to outperform state-of-the-art, manual-feature-engineering-based shallow classification. However, both deep and shallow machine learning algorithms have widely been applied to English Corpora, with little work being carried out to develop deep learning models for Persian sentiment analysis [33–42].

Persian is the official language of Iran and Afghanistan, with more than one hundred million speakers. A large amount of unstructured text is available online, including newspapers, books, web pages, movie reviews, etc. [43–45]. In this study, a novel corpus for Persian sentiment analysis is developed and evaluated using both shallow and deep machine learning algorithms. For shallow learning, logistic regression, support vector machine (SVM), and multilayer perceptron (MLP) classifiers are used. For deep learning, 1D convolutional neural network (CNN), 2D-CNN, stacked long short-term memory (LSTM), and Bidirectional LSTM algorithms have been utilized.

To the best of our knowledge, this is the first work on exploiting deep-learning-based automated feature engineering for Persian sentiment analysis. In addition, the *fastText* word embedding is used to obtain a vector representation of Persian words. The contextual features are extracted using deep learning algorithms and their polarity detection performances are compared with traditional shallow classifiers.

In summary, the paper reports two major contributions, outlined below:

1. We proposed novel architectures for deep learning classifiers such as convolutional neural network (CNN) and long short-term memory (LSTM) to identify the polarity of the Persian text;
2. An ablation study of our proposed deep learning reveals the importance of individual layers in the context of the complete framework.

The rest of the paper is organized as follows: In Section 2, related work on sentiment analysis for English, Persian, and other languages is presented. Section 3 presents our proposed novel approach for Persian sentiment analysis. In Section 4, experimental results are presented. Finally, Section 5 concludes this work with suggestions for future work.

## 2. Related Work

### 2.1. English Language

Recently, Pitsilis et al. [46] proposed a framework to discern hateful content in social media using a Recurrent Neural Network (RNN) classifier. The authors collected data from Twitter and fed the extracted features into an RNN classifier. The results demonstrated the effectiveness of RNN (achieving up to 95.33%) as compared to a state-of-the-art machine learning algorithm such as SVM (75.22%). Chen et al. [47] proposed a novel framework to improve sentence-level sentiment analysis by employing Long-Short-Term Memory with a conditional random field layer (BiLSTM-CRF). The comparative simulation results with benchmark datasets showed that their proposed framework improved the overall accuracy of sentence-level sentiment analysis. Hassan et al. [48] proposed an approach to detect polarity using CNN and LSTM using pre-trained word vectors of IMDB movie reviews. The CNN classifier developed in this approach consisted of two convolutional layers and two pooling layers. The experimental results showed that the combined CNN and LSTM model achieved up to 88.3% accuracy and outperformed CNN (87%) and LSTM (81.8%) models. Similarly, Shen et al. [49] proposed a novel approach to combine CNN and bidirectional Long-Short-Term Memory (BLSTM) to detect polarity for movie reviews. The combined CNN-LSTM model outperformed CNN and LSTM classifiers by achieving an accuracy of up to 89.7% as compared to 83.9% and 78.4%, respectively.

Nguyen et al. [50] proposed a novel method to detect polarity in news articles using a deep learning classifier. The authors used different websites to collect more than one million news articles and fed the preprocessed embedded vectors into CNN, LSTM and convolutional LSTM (CLSTM). The comparative simulation results showed that the CLSTM model outperformed CNN and LSTM classifiers by achieving an accuracy of up to 96.52%

as compared to the 92.3% and 91.19% accuracy achieved by CNN and LSTM, respectively. Liao et al. [51] proposed a framework to understand users' satisfaction with a product using a deep learning classifier. The authors used the Twitter API to collect the tweets. The CNN classifier developed in this framework contained one convolutional layer and one pooling layer. Their experimental results showed that CNN outperformed SVM and Naive Bayes classifiers by achieving the accuracy up to 95% as compared to the 70% and 62.25% accuracy achieved by SVM and Naive Bayes, respectively. Zhao et al. [52] proposed a model to detect sentiment in product reviews using a deep learning classifier. The words are first converted into an embedding vector representation, and then a deep learning classifier is used to classify sentiments. The experimental results showed the effectiveness of their proposed model, where CNN achieved up to 87% accuracy as compared to LSTM (82%). Ouyang et al. [53] proposed a framework to detect polarity in English movie reviews using a deep learning classifier. The movie reviews were collected from rottentomatoes.com. The dataset consisted of five labels: positive, somewhat positive, neutral, somewhat negative and negative. The experimental results showed that the RNN achieved an accuracy of up to 78.34% as compared to SVM (62%).

## 2.2. Persian Language

Basiri et al. [54] proposed a lexicon for Persian that consisted of 150,000 sentences along with labels indicating the polarity of the sentences. The authors used a Naive Bayes classifier to evaluate the performance of their proposed approach and achieved 91% accuracy. Ebrahimi et al. [55] proposed a supervised method to develop a sentiment lexicon for the Persian language by extracting features in movie reviews. The authors used an SVM classifier to evaluate the performance of the approach and achieved 80% accuracy. Saraee et al. [56] proposed a novel approach to detect polarity in Persian movie reviews using n-gram features. Their proposed approach consisted of stemming and feature selection. The authors used a Naive Bayes classifier to evaluate the performance and achieved 82.26% accuracy.

Amiri et al. [57] developed a lexicon to detect polarity for multi-domain product and movie reviews in Persian. The authors collected sentences from people's communication and then pre-processed them. The authors used an SVM classifier to evaluate the performance of the lexicon and achieved an accuracy of 69%. Vaziripour et al. [58] proposed an approach to automatically classify the sentiment in Persian tweets. The experimental results showed that SVM achieved up to 70% accuracy as compared to Naïve Bayes (56%). Sabeti et al. [59] developed a lexicon for Persian. The lexicon achieved an accuracy of 81.06% using K-nearest neighbors (KNN).

Basiri et al. [60] proposed a method to detect polarity in Persian product reviews. The experimental results showed that SVM achieved an accuracy up to 68% and outperformed Naive Bayes (62%) and MLP (60%) classifiers. Alimardani et al. [61] proposed an approach to detect polarity in Persian hotel reviews. The developed lexicon contained more than one thousand words along with their polarity. The proposed approach collected hotel reviews and extracted part-of-speech tag features. The performance analysis revealed that logistic regression outperformed the SVM classifier by achieving an accuracy up to 85.9% as compared to the 82.4% accuracy achieved by SVM.

## 2.3. Other Languages

Day et al. [62] proposed a framework to explore the impact of deep learning for sentiment analysis on Google App mobile reviews in Chinese. CNN outperformed SVM and Naive Bayes classifiers by achieving an accuracy up to 94% as compared to the 76.46% and 74.12% accuracy achieved by SVM and Naive Bayes, respectively. Valverde et al. [63] proposed a novel approach to detect polarity for Spanish product services using a deep learning classifier. The product reviews were manually collected from e-commerce websites. After the pre-processing, the word embeddings was used to obtain the vectors of the words, followed by the CNN classifier. The authors achieved 88.7% accuracy.

Baniata et al. [64] proposed a sentiment analysis method to identify the polarity for Arabic text using a deep learning classifier. The combination of CNN and LSTM classifiers, trained on Arabic reviews, achieved a high accuracy of 86.43% in comparison to CNN (66.26%) and LSTM (65.34%). Xiao et al. [65] proposed an approach to detect polarity in Chinese text. The authors used 1D and 2D CNN classifiers and achieved the highest accuracy, of 93.4%, with 2D CNN. Furthermore, the experimental result showed that the character-level approach outperformed word embedding for Chinese words.

Dahou et al. [66] proposed an approach to detect polarity for Arabic text. The authors used a web crawler to build a corpus and trained word embedding to represent words in the corpus. The presented results showed that a CNN classifier with Arabic reviews outperformed other state-of-the-art supervised learning algorithms, achieving 89.7% as compared to SVM (72.4%) and Naïve Bayes (69.5%). Le et al. [67] proposed an approach for sentiment analysis for the Indonesian language to detect the polarity in the sentences or documents. The authors used four thousand movie reviews, labelled manually. Shaung et al. [68] proposed an approach using CNN-LSTM to detect polarity in English and Chinese product reviews. Their proposed CNN-LSTM model outperformed individual CNN and LSTM classifiers and achieved 81.86% accuracy. Wehrmann et al. [69] proposed a method to detect polarity in four different languages. However, their proposed method used machine translation to translate English tweets into German, Portuguese and Spanish. The experimental results showed the effectiveness of their proposed model and achieved up to 76.2% accuracy with CNN, as compared to LSTM, which achieved 64.7% accuracy. More recently, Sankar et al. [70] proposed a novel framework, which uses training and test data from different datasets, such as IMDB and Rotten Tomatoes, and applies deep learning classifiers to identify the polarity of the sentence. In other hand, Ali et al. [71] implemented a sentiment classification approach using deep learning algorithms such as LSMT and CNN and hybrid CNN and LSTM to identify the polarity of the sentence. However, most of the approaches are not suitable to determine the polarity for Persian sentences.

### 3. Methodology

In this section, the proposed approach for Persian movie reviews is discussed in detail. Figure 1 depicts the proposed framework, and details are presented in subsequent sections.

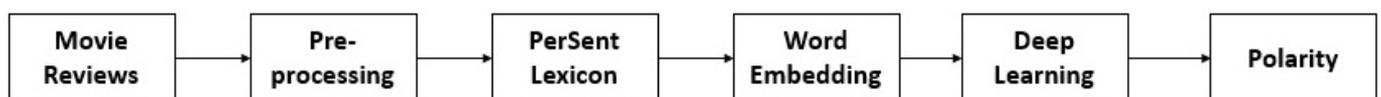


Figure 1. Proposed Framework.

#### 3.1. Data Pre-Processing

The novel dataset used in this work is collected manually from Persian movie websites such as [www.caffecinema.com](http://www.caffecinema.com) (accessed on 15 March 2021) and [www.cinematicket.org](http://www.cinematicket.org) (accessed on 15 March 2021). A subset of the dataset was used to train the neural network (60% training dataset), and the rest of the data (40%) were used to test and validate the performance of the trained neural network testing set (30%), and validation set (10%). There were two types of label in the dataset: positive and negative. The reviews were manually annotated by three native Persian speakers aged between 30 and 50 years old. The reviewers received a degree in Persian from the University of Tehran. There were 1021 positive and 989 negative reviews. After data collection, the dataset was pre-processed using tokenisation and normalisation techniques. The process of breaking a sentence into words is called tokenisation. In order to tokenise Persian sentences, the NLTK tokeniser is used. For example, “The movie is great” is tokenised into “The”, “movie”, “is”, and “great”. The normalisation technique is used to replace abbreviations and replace them with their actual meaning. In addition, normalisation is used to replace words with their actual meaning. For example, “I am happyyyyyyy” is normalised to “I am happy”. A stemmer is used to change the words into their roots. For example, “fishing” or “fisher” is stemmed

into “fish”. In order to do the stemming process, a Persian Python tool called HAZM is used. HAZM contains text cleaning, sentence word tokenizer, word lemmatizer, part-of-speech tagger, shallow parser and dependency parser [72–76]. Figure 2 shows examples of Persian sentences.

Positive	Negative
فیلم بینظیری بود بعد از چند سال واقعا بهترین فیلمی بود که دیدم ارزش چند بار دیدنو داره	ظاهراً کارگردان فقط قصد داشته با یک موضوع جنجالی و کنجکاو برانگیزجذب مخاطب کند بدون اینکه فیلمش پتانسیل لازم برای پاسخ به مخاطب را در این ژانر داشته باشد
پیشنهاد میکنم حتما ببینید بازیها خیلی طبیعی و خوب بود	بازی ها کاملاً متوسط است و شاهد سکانس بد هستیم
فیلم از همه لحاظ کارگردانی فیلم نامه بازی و بسیار خوب بود واقعا بدون رقیب بود	نا امید شدم خیلی خسته کننده بود طوری که درنهایت مردم فریاد زدند ا خ جووون بالاخره تموم شد

Figure 2. Persian Sentences Examples.

N-gram features are extracted from text. The n-gram size one is called unigram, size two is bigram and size three is trigram [77]. The n-gram features are extracted from Persian movie reviews. HAZM reads the text and assigns POS tag into words [78]. These tags are noun, verb, adjective, adverbs and etc. The HAZM tool is used to extract POS tag (Adjective, Adverb, Verb and Noun) from Persian movie reviews. After extracting n-grams and POS tag features, the PerSent lexicon is used to assign polarity into features. The lexicon contains 1500 Persian words along with their part-of-speech tag (POS) and the polarity of the words [21].

**Word embedding:** In order to train the CNN classifier, the Persian tokenized words are converted into a three hundred dimensional vector using fastText pretrained embedding Python package. Figure 3 shows the process of word embedding with the Persian sentence example, <mn iyn fylm xlyl dwst dArm>.

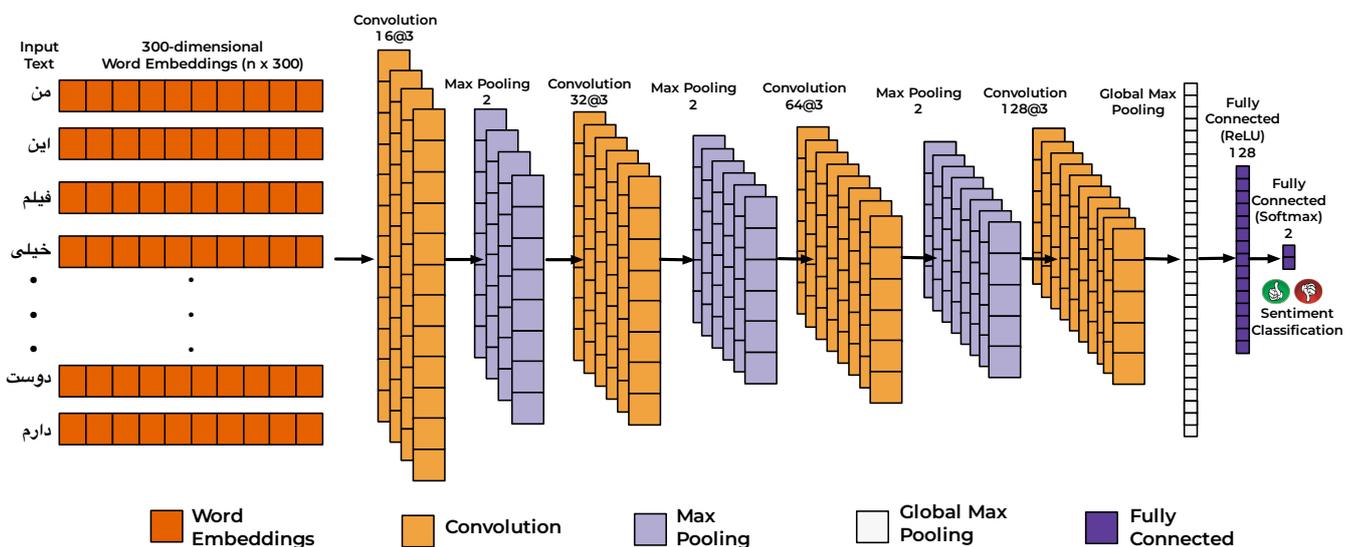


Figure 3. CNN Classifier.

### 3.2. Classification

#### Convolutional Neural Network

The developed CNN classifier is shown in Figure 3. The CNN classifier consists of input, output, and hidden layers. The hidden layers constitute of convolutional, pooling, fully connected and normalization layers [33,79,80]. In experiments, the best results were obtained using an 11-layered CNN architecture, as shown in Figure 3. In the first convolu-

tional layers, 16 feature maps with a kernel size of 2 are used. In the second convolutional layer, 32 feature maps with a kernel size 2 are used. The layers are followed by a max pooling layer of size 2. In the fourth convolutional layer, 64 feature maps with a kernel size 2 are used. The final convolutional layer is followed by a max pooling with window size 2. The feature extraction framework is followed by fully connected layers of size 5000, 500 and 2.

**Long Short Term Memory (LSTM):** LSTM is a successful augmented recurrent neural network model which is used to learn sequential information with dependencies that LSTM can store and use to compute information for a long period of time. LSTM was originally proposed in [81] by Hochreiter et al. The developed LSTM network, presented in Figure 4, consists of an input layer, two stacked LSTM layers, and a fully connected layer. Specifically, the network consists of two bidirectional LSTM layers followed by two dropout layers, one dense layer and one activation function layer. The fastText word embedding is fed into stacked LSTM layers. The output of the second LSTM layer was then fed into the fully connected (dense) layer with two neurons. The architecture was trained using an Adam optimizer with a dropout probability of 0.2. The main reason for employing LSTM is its recurrent nature and ability to help the model with classification data. It has to be noted that previous works, such as [82–85], reveal that LSTM, due to its inherent recurrent nature, can have better model long-term dependencies and exploit temporal correlation in inputs, as compared to the MLP.

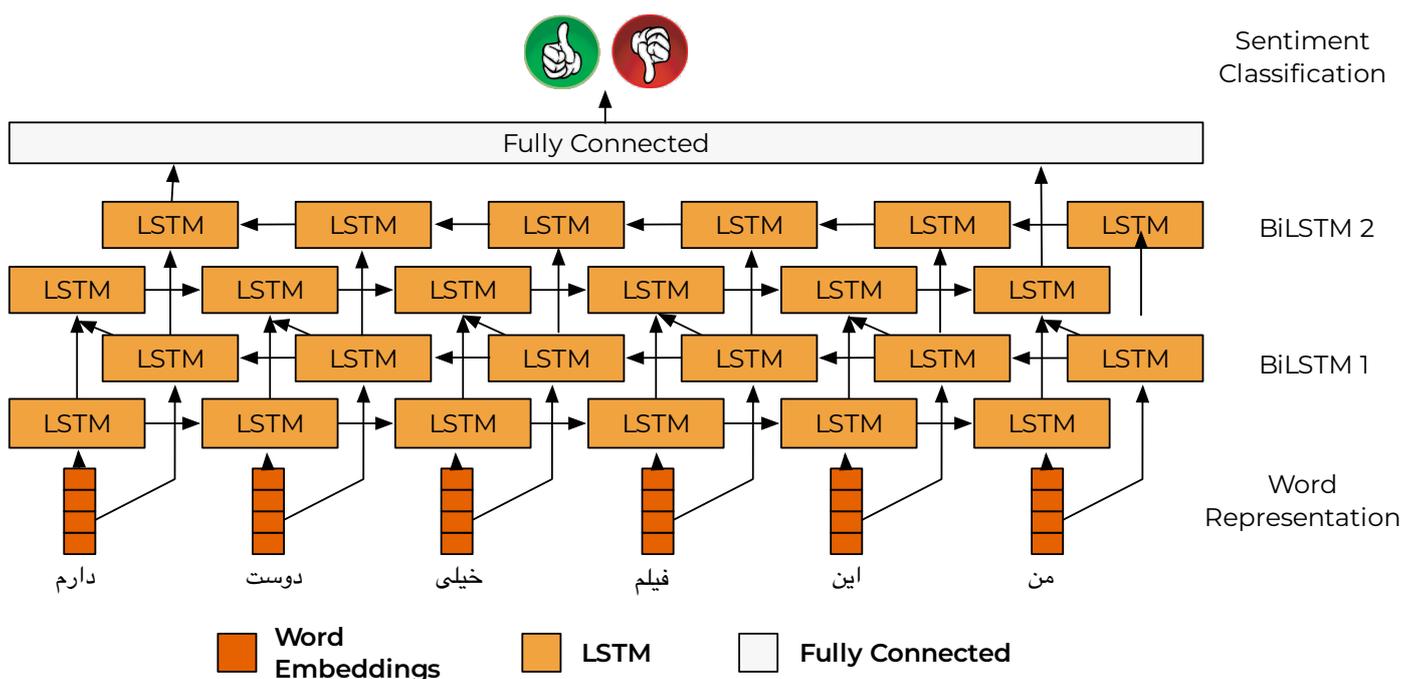


Figure 4. Long Term Short Memory.

**MultiLayer perceptron (MLP):** MLPs are made up of highly interconnected processing elements called neurons, processing information by their state response and learning from examples. A neuron in an MLP is connected to several inputs with different associated weights. The output of a neuron is the summation of all connected inputs, followed by a non-linear processing unit, called a transfer function. The main objective of MLP is to transform the inputs into meaningful outputs by learning the input–output relationship, and offering viable solutions to unseen problems (a generalization capability). Therefore, the capacity to learn from examples is one of the most desirable features of neural network models. The goal of training is to learn the desired system behavior and adjust the network parameters (interconnections weights) to map (learn) the input–output relationship and

minimize the cost function. The processed embedding vectors were fed into an MLP model with 1 or 2 hidden layers and 10 to 150 hidden neurons per layer.

**Autoencoder:** An autoencoder (AE) is a type of unsupervised learning algorithm, typically used for dimensionality reduction purposes. The AE standard configuration includes one input layer, one output layer and one hidden layer. It compresses the input data  $x$  into a lower dimension  $h$  through the encoding process:

$$h = g(xw + b) \quad (1)$$

where  $x$ ,  $w$ ,  $b$  are the input vector, weight matrix, the bias vector, respectively and  $g$  is the activation function. Then, it attempts to reconstruct the same set of input ( $x$ ) from the compressed representation ( $h$ ) through the decoding process:

$$\tilde{x} = g(hw^T + b) \quad (2)$$

#### 4. Experimental Results

In this section, we describe the experimental setup, followed by results and discussions. To evaluate the performance of the proposed approach, movie reviews are used. For data labelling, the PerSent lexicon is used, which assigns a polarity to individual events in the dataset. The trained and test dataset is converted into vectors to train LSTM, CNN, and SVM classifiers. The n-gram features (bigram, trigram) and POS features (noun, adjective, verb and adverb) are extracted from Persian movie reviews. The extracted features are converted into bag of words and principal component analysis is used to reduce the dimensionality of data to two hundred dimensions. The extracted features are fed into an SVM and a logistic regression classifier.

The parameters of LR, SVM, LSTM, and CNN models are as follows: word embedding dimensions are 300, the number of epochs equal to 200, and batch size is 128. The ML classifiers are trained to classify the sentences into either positive or negative. In addition, it has been shown that the size of the filter has a positive impact on the final results and model achieved a better performance when the filter size was set to a smaller number such as 3.

Table 1 presents the results of SVM using various features and combination of the features. The results shows that nouns outperformed other features. For testing with LSTM and CNN, movie reviews are converted into three hundred dimensional vectors using fasttext. Table 3 presents the results of CNN and LSTM classifiers and a comparison with MLP and autoencoder. The experimental results demonstrate the effectiveness of the autoencoder as compared to MLP. To evaluate the performance of proposed approaches, precision, recall, f-Measure, and prediction accuracy are used as performance matrices

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F\_measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

where TP represents, TN, FP, and FN represents true positive, true negative, false positive, and false negative, respectively.

**Table 1.** Deep Learning Classifiers Results on Persian Movie reviews.

Classifier	Precision	Recall	F-Measure	Accuracy (%)
Ghasemi et al. [86]	0.63	0.62	0.63	63.94
Asli et al. [87]	0.68	0.67	0.68	68.23
Amiri et al. [57]	0.69	0.68	0.69	70.01
Basiri et al. [54]	0.72	0.71	0.72	72.81
MLP	0.78	0.78	0.78	78.49
MLP-Autoencoder	0.8	0.8	0.8	80.08
2D-CNN	0.82	0.82	0.82	82.47
1D-CNN	0.84	0.83	0.83	82.86
Minimal-RNN	0.86	0.86	0.86	85.64
Stacked-LSTM	0.94	0.94	0.94	93.65
<b>Stacked-BiLSTM</b>	<b>0.96</b>	<b>0.96</b>	<b>0.96</b>	<b>95.61</b>

Table 2 indicates the results for SVM and logistic regression. The experimental results shows that the noun achieved better performance as compared to other features such as adjective, adverb, verb, bigram and trigram. The experimental results demonstrate that the LR achieved better accuracy, precision, recall and f-measure as compared to SVM.

**Table 2.** SVM vs. Logistic Regression (LR) on Persian Movie reviews.

Feature	Precision		Recall		F-Measure		Accuracy (%)	
	SVM	LR	SVM	LR	SVM	LR	SVM	LR
Bigram	0.65	0.66	0.61	0.67	0.53	0.65	61.27	66.59
Trigram	0.81	0.88	0.75	0.88	0.71	0.88	74.53	87.96
Adjective	0.72	0.87	0.85	0.85	0.78	0.79	84.57	85.07
Adverb	0.76	0.81	0.87	0.86	0.81	0.82	86.95	85.88
<b>Noun</b>	<b>0.79</b>	<b>0.91</b>	<b>0.89</b>	<b>0.89</b>	<b>0.84</b>	<b>0.85</b>	<b>88.93</b>	<b>89.37</b>
Verb	0.78	0.78	0.88	0.88	0.83	0.83	88.4	88.4

The experimental results indicated that using deep learning approaches enhances the performance of sentiment analysis in Persian dataset. The proposed model does not required any feature engineering to extract special features such as n-gram, POS and concept extraction like previous approaches. It is to be noted that the deep learning approaches are based on the pre-trained word vector representation; despite the complexity of the Persian language and the simplicity of the proposed model, there are significant improvements in the BiLSTM in terms of the F-measure and the accuracy compared to traditional machine learning classifiers.

On the other hand, deep learning classifiers achieved a better performance, such as the BiLSTM and 2D-CNN, with fastText word embedding used as features. The fastText model consists of high-quality vector representation using the semantic and syntactical information from the texts; it can also cover the out-of-vocabulary words.

The deep learning approaches are black-box; we cannot understand how the sentiment polarity assigned to sentences, for example, the polarity shifters such as negation, can change the overall polarity of the sentences. However, the Persian language consists of rhetorical and sarcastic sentences which cannot be detected by black-box deep learning approaches.

The deep learning architectures trained and validated using TensorFlow library and NVIDIA Titan X GPU. The deep learning models were trained for 200 epochs using back propagation, with the Adam optimizer minimizing the categorical cross entropy loss function.

Table 3 presents the results of SVM and LR using various features on hotel reviews. Experimental results indicate that the LR achieved better accuracy as compared with SVM. The results show that the use of adjective outperformed other features. Additionally, the adverb feature achieved lower accuracy as compared with other features.

**Table 3.** SVM vs. LR on Hotel reviews dataset.

Feature	Precision		Recall		F-Measure		Accuracy (%)	
	SVM	LR	SVM	LR	SVM	LR	SVM	LR
Bigram	0.71	0.72	0.71	0.72	0.71	0.72	71.25	72
Trigram	0.73	0.74	0.73	0.74	0.73	0.74	73.5	74
<b>Adjective</b>	0.75	0.76	0.78	0.79	0.76	0.77	76.24	77.06
Adverb	0.63	0.64	0.62	0.63	0.62	0.62	62.72	62.78
Noun	0.67	0.69	0.68	0.70	0.67	0.69	68.21	69.08
Verb	0.68	0.69	0.69	0.71	0.68	0.69	68.54	69.97

Furthermore, Table 4 presents the results of CNN and LSTM classifiers on the hotel reviews dataset. Experimental results show the 2D-CNN achieved better accuracy as compared with other classifiers. Additionally, the stacked-BiLSTM received a lower performance as compared with other classifiers; however, it achieved better precision as compared with other classifiers.

**Table 4.** Deep Learning Classifiers Results on Hotel reviews dataset.

Classifier	Precision	Recall	F-Measure	Accuracy (%)
Stacked-BiLSTM	0.53	0.91	0.67	74.49
1D-CNN	0.73	0.80	0.76	78.02
Stacked-LSTM	0.79	0.81	0.80	81.04
<b>2D-CNN</b>	0.89	0.89	0.89	89.76

Table 5 presents the comparison results for the proposed 1D-CNN with different layers. Experimental results show that the five-layered 1D-CNN architecture generally outperforms other layered architecture. The time to train each model is shown in the last column of Table 5. Additionally, the recall and F-measure for five-layers outperformed other architectures compared with 2-, 3-, 4-, and 6-layered models.

**Table 5.** Comparison of 1D-CNN Layers on Persian Movie reviews.

Layer	Precision	Recall	F-Measure	Accuracy (%)	Time
2	0.72	0.73	0.72	73.55	2 m 31 s
3	0.74	0.74	0.74	74.26	3 m 12 s
4	0.81	0.79	0.79	79.01	3 m 44 s
<b>5</b>	0.84	0.83	0.83	82.86	4 m 22 s
6	0.78	0.76	0.76	78.42	5 m 24 s

Table 6 presents the comparison results for the proposed 2D-CNN with different layers. Experimental results show that the five-layer CNN generally outperforms other layers. The time to train each model is shown in the last column of Table 6. Additionally, the recall and F-measure for layer 5 outperformed other layers.

**Table 6.** Comparison of 2D-CNN Layers on Persian Movie reviews.

Layer	Precision	Recall	F-Measure	Accuracy (%)	Time
2	0.72	0.70	0.70	74.05	2 m 21 s
3	0.73	0.71	0.71	74.22	3 m 23 s
4	0.78	0.78	0.78	78.26	5 m 5 s
5	0.82	0.82	0.82	82.47	6 m 33 s
6	0.75	0.76	0.75	76.51	8 m 25 s

Table 7 presents the comparison results for different layers of LSTM. Experimental results show that the two-layer LSTM achieved better results as compared with other layers. The time to train the model is shown in Table 7. Additionally, experimental results show that the two-layer LSTM achieved better precision, recall, f-measure as compared with other layers.

**Table 7.** Comparison of LSTM Layers on Persian Movie reviews.

Layer	Precision	Recall	F-Measure	Accuracy (%)	Time
2	0.94	0.94	0.94	93.65	6 m 49 s
3	0.84	0.83	0.83	84.26	7 m 28 s

Table 8 presents the comparison results for different layers of BiLSTM. Experimental results show that the two-layer BiLSTM achieved better results as compared with the one-layer BiLSTM. Additionally, experimental results show that the two-layer BiLSTM achieved better precision, recall and f-measure as compared with the one-layer BiLSTM. Experimental results show the training time is increased as the number of layers increases.

**Table 8.** Comparison of BiLSTM Layers on Persian Movie reviews.

Layer	Precision	Recall	F-Measure	Accuracy (%)	Time
2	0.96	0.96	0.96	95.61	7 m 23 s
3	0.84	0.86	0.85	85.05	9 m 24 s

The results demonstrate that the stacked-BiLSTM model outperformed all other methods. In addition, it is observed that CNN and LSTM classifiers can effectively detect the polarity of the movie reviews. Furthermore, they also help in detecting contextual information as compared to traditional classifiers. It is shown that deep learning approaches are more optimal for sentimental analysis due to their having less over-fitting and better generalization.

The hotel reviews corpus is used to compare how our approach performs in a new domain compared to state-of-the-art approaches, including multilingual methods. The hotel reviews were collected and, after data pre-processing, the extracted features were converted into TF-IDF and the machine learning was applied to evaluate the performance of the approach. The overall accuracy of their proposed approach is 87%. However, for their experiments, a five-fold cross-validation was used [61]. In our experiments, we used 60% for training, and the rest of the data were used for testing and to validate the performance of the trained classifiers, with 30% used for testing and 10% for the validation set.

Figure 5 presents the accuracy of both shallow and deep learning classifiers. It can be seen that the stacked-bidirectional-LSTM outperformed all other machine learning algorithms.

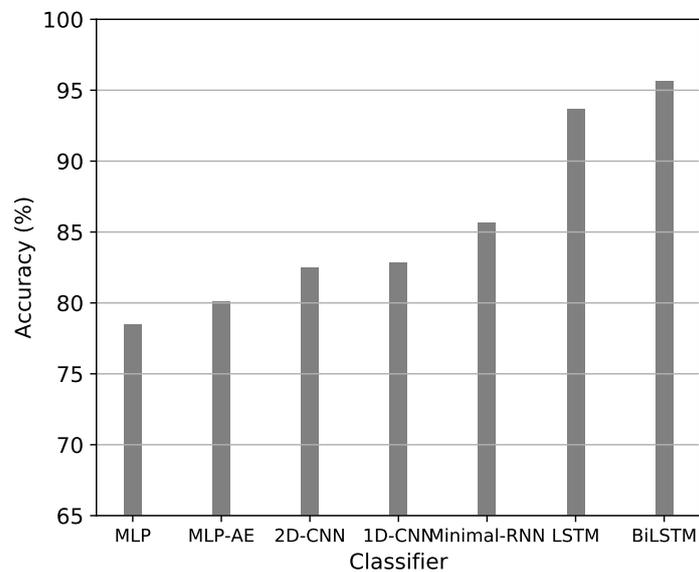


Figure 5. Deep Learning Classifiers' Accuracy on Movie reviews.

Figure 6 depicts the classification performance of SVM and LR using different feature types. It can be seen that the noun features help in achieving a better performance as compared to other features.

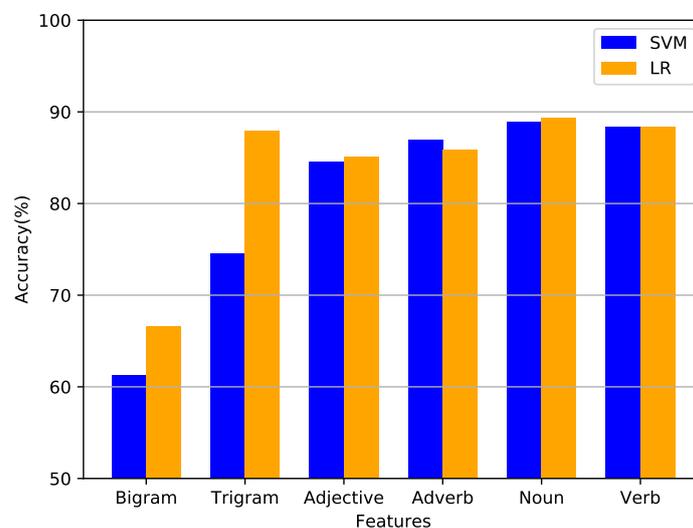


Figure 6. SVM vs. LR Classifier Accuracy on Movie reviews.

The results demonstrate that the stacked-BiLSTM model outperformed all other methods. In addition, it is observed that the CNN and LSTM classifiers can effectively detect the polarity of movie reviews. Furthermore, they also help in detecting contextual information as compared to traditional classifiers. It is shown that the deep learning approaches are more optimal for sentimental analysis due to their having less over-fitting and better generalization.

The LSTM achieved a better performance than shallow learning because, in deep learning approaches, there is no requirement for feature selection. The experimental results show that the LSTM can efficiently reflect the delays and make the input more convenient, and helps the LSTM achieve a better performance by optimizing the input form and

network structure based on the clearer physical meaning. However, in order to improve the performance further, there is a need to develop an additional lexicon or grammar-rule based approach to detect sarcastic and irony sentences' polarity. For example <cy bgm, nmydwnm xwb bwd yA bd bwd>, "What to say, I do not know the movie was good or bad". The experimental results indicated that the word embedding models can identify the overall sentiment polarity of the sentiment effectively. The classification performance is better for movie reviews than casual comments. This shows the feasibility of an LSTM-based approach for Persian movie sentiment classification.

The trigram performed well as compared to the bigram, as it is more informative; for example, it consists of three word combinations, making it easier for the machine learning model to identify the overall sentiment polarity. For example, <fyilm xwb bwd> "movie was great" as opposite to <fyilm> "movie". However, other studies, such as [88], showed that trigrams achieved a better performance as compared to bigrams.

## 5. Conclusions

With the rise of social media, people around the world openly share their opinions on different topics, movies, products, politics, etc. Sentiment analysis is widely used to automatically classify these sentiments as positive or negative. In the literature, much research has been conducted with English corpora but very little effort has been devoted to Persian sentiment analysis. In this study, a novel Persian architecture is developed for deep learning approaches. In order to evaluate the performance of the approach, the Persian movie reviews and hotel review dataset are used to test and validated using both shallow and deep learning algorithms. The simulation results demonstrate that deep learning outperformed state-of-the-art, shallow machine learning approaches. Specifically, the stacked-bidirectional-LSTM achieved the highest accuracy, of up to 95.61%, for the movie dataset, but not for the hotel dataset, where 2D-CNN did better (89.76%). Furthermore, we have evaluated the best performance of deep learning classifiers with different layers. Ongoing and future work aims to extend the developed corpus to multilingual product reviews.

**Author Contributions:** Conceptualization, K.D.; funding acquisition, K.D.; investigation, K.D. and A.A.; methodology, K.D and M.G.; project administration, K.D., M.G., A.A. and A.H.; resources, K.D., Mandar Gogate, A.A. and A.H.; supervision, H.L. and A.H.; writing—original draft, M.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** Amir Hussain was supported by the UK Engineering and Physical Sciences Research Council (EPSRC)-Grants Ref. EP/M026981/1, EP/T021063/1, EP/T024917/1.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Cambria, E.; Poria, S.; Hazarika, D.; Kwok, K. SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. *AAAI* **2018**, *32*, 1795–1802. Available online: <https://sentit.net/senticnet-5.pdf> (accessed on 15 March 2021).
2. Dashtipour, K.; Poria, S.; Hussain, A.; Cambria, E.; Hawalah, A.Y.; Gelbukh, A.; Zhou, Q. Multilingual sentiment analysis: state of the art and independent comparison of techniques. *Cogn. Comput.* **2016**, *8*, 757–771. [[CrossRef](#)] [[PubMed](#)]
3. Alqarafi, A.S.; Adeel, A.; Gogate, M.; Dashitpour, K.; Hussain, A.; Durrani, T. Toward's Arabic multi-modal sentiment analysis. In *International Conference in Communications, Signal Processing, and Systems*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 2378–2386.
4. Gepperth, A.R.; Hecht, T.; Gogate, M. A generative learning approach to sensor fusion and change detection. *Cogn. Comput.* **2016**, *8*, 806–817. [[CrossRef](#)]
5. Suresh Kumar, S.; Dashtipour, K.; Abbasi, Q.H.; Imran, M.A.; Ahmad, W. A review on wearable and contactless sensing for COVID-19 with policy challenges. *Front. Commun. Netw.* **2021**. [[CrossRef](#)]

6. Taylor, W.; Shah, S.A.; Dashtipour, K.; Zahid, A.; Abbasi, Q.H.; Imran, M.A. An intelligent non-invasive real-time human activity recognition system for next-generation healthcare. *Sensors* **2020**, *20*, 2653. [CrossRef]
7. Adeel, A.; Gogate, M.; Farooq, S.; Ieracitano, C.; Dashtipour, K.; Larijani, H.; Hussain, A. A survey on the role of wireless sensor networks and IoT in disaster management. In *Geological Disaster Monitoring Based on Sensor Networks*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 57–66.
8. Gogate, M.; Hussain, A.; Huang, K. Random Features and Random Neurons for Brain-Inspired Big Data Analytics. In Proceedings of the 2019 International Conference on Data Mining Workshops (ICDMW), Beijing, China, 8–11 November 2019; pp. 522–529.
9. Gogate, M.; Adeel, A.; Hussain, A. A novel brain-inspired compression-based optimised multimodal fusion for emotion recognition. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 November–1 December 2017; pp. 1–7.
10. Wang, Y.; Wang, S.; Tang, J.; Liu, H.; Li, B. *Unsupervised Sentiment Analysis for Social Media Images*; IJCAI: Tempe, AZ, USA, 2015; pp. 2378–2379.
11. Liu, B. Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* **2012**, *5*, 1–167. [CrossRef]
12. Liu, B. Sentiment Analysis and Subjectivity. *Handb. Nat. Lang. Process.* **2010**, *2*, 627–666.
13. Jiang, F.; Kong, B.; Li, J.; Dashtipour, K.; Gogate, M. Robust visual saliency optimization based on bidirectional Markov chains. *Cogn. Comput.* **2020**, 1–12. [CrossRef]
14. Ahmed, R.; Gogate, M.; Tahir, A.; Dashtipour, K.; Al-tamimi, B.; Hawalah, A.; El-Affendi, M.A.; Hussain, A. Deep Neural Network-Based Contextual Recognition of Arabic Handwritten Scripts. *Entropy* **2021**, *23*, 340. [CrossRef]
15. Gogate, M.; Adeel, A.; Hussain, A. Deep learning driven multimodal fusion for automated deception detection. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 November–1 December 2017; pp. 1–6.
16. Nisar, S.; Tariq, M.; Adeel, A.; Gogate, M.; Hussain, A. Cognitively inspired feature extraction and speech recognition for automated hearing loss testing. *Cogn. Comput.* **2019**, *11*, 489–502. [CrossRef]
17. Zhou, R.; Niu, L. Feature Selection of Network Data VIA  $l_2$ , p Regularization. *Cogn. Comput.* **2020**, *12*, 1217–1232. [CrossRef]
18. Fei, X.; Shen, L.; Ying, S.; Cai, Y.; Zhang, Q.; Kong, W.; Zhou, W.; Shi, J. Parameter Transfer Deep Neural Network for Single-Modal B-Mode Ultrasound-Based Computer-Aided Diagnosis. *Cogn. Comput.* **2020**, *12*, 1252–1264. [CrossRef]
19. Rajabi, Z.; Valavi, M.R.; Hourali, M. A Context-Based Disambiguation Model for Sentiment Concepts Using a Bag-of-Concepts Approach. *Cogn. Comput.* **2020**, *12*, 1299–1312. [CrossRef]
20. Zahid, A.; Dashtipour, K.; Carranza, I.E.; Abbas, H.; Ren, A.; Cumming, D.R.; Grant, J.P.; Imran, M.A.; Abbasi, Q.H. Machine Learning Framework for the Detection of Anomalies in Aqueous Solutions Using Terahertz Waves. *Res. Sq.* **2021**. [CrossRef]
21. Dashtipour, K.; Hussain, A.; Zhou, Q.; Gelbukh, A.; Hawalah, A.Y.; Cambria, E. PerSent: A freely available Persian sentiment lexicon. In *International Conference on Brain Inspired Cognitive Systems*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 310–320.
22. Dashtipour, K.; Gogate, M.; Adeel, A.; Algarafi, A.; Howard, N.; Hussain, A. Persian named entity recognition. In Proceedings of the 2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI\*CC), Oxford, UK, 26–28 July 2017; pp. 79–83.
23. Dashtipour, K.; Gogate, M.; Li, J.; Jiang, F.; Kong, B.; Hussain, A. A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks. *Neurocomputing* **2020**, *380*, 1–10. [CrossRef]
24. Dashtipour, K.; Gogate, M.; Cambria, E.; Hussain, A. A Novel Context-Aware Multimodal Framework for Persian Sentiment Analysis. *arXiv* **2021**, arXiv:2103.02636.
25. Zheng, Y.; Zhao, X.; Yao, L. Mixture Kernel Density Estimation and Remedied Correlation Matrix on the EEG-Based Copula Model for the Assessment of Visual Discomfort. *Cogn. Comput.* **2020**, *12*, 1130–1143. [CrossRef]
26. Liaquat, S.; Dashtipour, K.; Zahid, A.; Assaleh, K.; Arshad, K.; Ramzan, N. Detection of atrial fibrillation using a machine learning approach. *Information* **2020**, *11*, 549. [CrossRef]
27. Asad, S.M.; Dashtipour, K.; Hussain, S.; Abbasi, Q.H.; Imran, M.A. Travelers-Tracing and Mobility Profiling Using Machine Learning in Railway Systems. In Proceedings of the 2020 International Conference on UK-China Emerging Technologies (UCET), Glasgow, UK, 20–21 August 2020; pp. 1–4.
28. Das, S.; Kalita, H.K. Sentiment Analysis for Web-based Big Data: A Survey. *Int. J. Adv. Res. Comput. Sci.* **2017**, *8*. Available online: <https://fardapaper.ir/mohavaha/uploads/2018/11/Fardapaper-Sentiment-Analysis-for-Web-based-Big-Data-A-Survey.pdf> (accessed on 15 March 2021)
29. Medhat, W.; Hassan, A.; Korashy, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Eng. J.* **2014**, *5*, 1093–1113. [CrossRef]
30. Seiti, H.; Hafezalkotob, A. A New Risk-Based Fuzzy Cognitive Model and Its Application to Decision-Making. *Cogn. Comput.* **2019**. [CrossRef]
31. Dashtipour, K.; Gogate, M.; Adeel, A.; Hussain, A.; Alqarafi, A.; Durrani, T. A comparative study of persian sentiment analysis based on different feature combinations. In *International Conference in Communications, Signal Processing, and Systems*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 2288–2294.
32. Yu, Z.; Machado, P.; Zahid, A.; Abdulghani, A.M.; Dashtipour, K.; Heidari, H.; Imran, M.A.; Abbasi, Q.H. Energy and performance trade-off optimization in heterogeneous computing via reinforcement learning. *Electronics* **2020**, *9*, 1812. [CrossRef]
33. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [CrossRef] [PubMed]

34. Shiva, A.S.; Gogate, M.; Howard, N.; Graham, B.; Hussain, A. Complex-valued computational model of hippocampal CA3 recurrent collaterals. In Proceedings of the 2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI\*CC), Oxford, UK, 26–28 July 2017; pp. 161–166.
35. Adeel, A.; Gogate, M.; Hussain, A.; Whitmer, W.M. Lip-reading driven deep learning approach for speech enhancement. *IEEE Trans. Emerg. Top. Comput. Intell.* **2019**. [[CrossRef](#)]
36. Dashtipour, K.; Hussain, A.; Gelbukh, A. Adaptation of sentiment analysis techniques to Persian language. In *International Conference on Computational Linguistics and Intelligent Text Processing*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 129–140.
37. Gasparini, S.; Campolo, M.; Ieracitano, C.; Mammone, N.; Ferlazzo, E.; Sueri, C.; Tripodi, G.G.; Aguglia, U.; Morabito, F.C. Information Theoretic-Based Interpretation of a Deep Neural Network Approach in Diagnosing Psychogenic Non-Epileptic Seizures. *Entropy* **2018**, *20*, 43. [[CrossRef](#)] [[PubMed](#)]
38. Dashtipour, K.; Ieracitano, C.; Morabito, F.C.; Raza, A.; Hussain, A. An Ensemble Based Classification Approach for Persian Sentiment Analysis. In *Progresses in Artificial Intelligence and Neural Systems*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 207–215.
39. Tan, L.; Li, C.; Huang, J. Neural Network-Based Event-Triggered Adaptive Control Algorithms for Uncertain Nonlinear Systems with Actuator Failures. *Cogn. Comput.* **2020**, *12*, 1370–1380. [[CrossRef](#)]
40. Adeel, A.; Gogate, M.; Hussain, A. Towards next-generation lipreading driven hearing-aids: A preliminary prototype demo. In Proceedings of the International Workshop on Challenges in Hearing Assistive Technology (CHAT-2017), Stockholm, Sweden, 19 August 2017.
41. Churher, A.; Ullah, R.; Ahmad, J.; ur Rehman, S.; Masood, F.; Gogate, M.; Alqahtani, F.; Nour, B.; Buchanan, W.J. An Experimental Analysis of Attack Classification Using Machine Learning in IoT Networks. *Sensors* **2021**, *21*, 446. [[CrossRef](#)] [[PubMed](#)]
42. Howard, N.; Adeel, A.; Gogate, M.; Hussain, A. Deep Cognitive Neural Network (DCNN). U.S. Patent App. 16/194,721, 23 May 2019.
43. Almeida, T.A.; Silva, T.P.; Santos, I.; Hidalgo, J.M.G. Text normalization and semantic indexing to enhance instant messaging and SMS spam filtering. *Knowl. Based Syst.* **2016**, *108*, 25–32. [[CrossRef](#)]
44. Dashtipour, K.; Raza, A.; Gelbukh, A.; Zhang, R.; Cambria, E.; Hussain, A. Persent 2.0: Persian sentiment lexicon enriched with domain-specific words. In *International Conference on Brain Inspired Cognitive Systems*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 497–509.
45. Jiang, F.; Dashtipour, K.; Hussain, A. A survey on deep learning for the routing layer of computer network. In Proceedings of the 2019 UK/China Emerging Technologies (UCET), Glasgow, UK, 21–22 August 2019; pp. 1–4.
46. Pitsilis, G.K.; Ramampiaro, H.; Langseth, H. Detecting Offensive Language in Tweets Using Deep Learning. *arXiv* **2018**, arXiv:1801.04433.
47. Du, H.; Xu, X.; Cheng, X.; Wu, D.; Liu, Y.; Yu, Z. Aspect-specific sentimental word embedding for sentiment analysis of online reviews. In Proceedings of the 25th International Conference Companion on World Wide Web, International World Wide Web Conferences Steering Committee, Montreal, QC, Canada, 11–15 April 2016; pp. 29–30.
48. Hassan, A.; Mahmood, A. Deep Learning approach for sentiment analysis of short texts. In Proceedings of the 2017 3rd International Conference on Control, Automation and Robotics (ICCAR), Nagoya, Japan, 24–26 April 2017; pp. 705–710.
49. Shen, Q.; Wang, Z.; Sun, Y. Sentiment Analysis of Movie Reviews Based on CNN-BLSTM. In *International Conference on Intelligence Science*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 164–171.
50. Nguyen, D.; Vo, K.; Pham, D.; Nguyen, M.; Quan, T. A deep architecture for sentiment analysis of news articles. In *International Conference on Computer Science, Applied Mathematics and Applications*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 129–140.
51. Liao, S.; Wang, J.; Yu, R.; Sato, K.; Cheng, Z. CNN for situations understanding based on sentiment analysis of twitter data. *Procedia Comput. Sci.* **2017**, *111*, 376–381. [[CrossRef](#)]
52. Zhao, W.; Guan, Z.; Chen, L.; He, X.; Cai, D.; Wang, B.; Wang, Q. Weakly-supervised deep embedding for product review sentiment analysis. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 185–197. [[CrossRef](#)]
53. Ouyang, X.; Zhou, P.; Li, C.H.; Liu, L. Sentiment analysis using convolutional neural network. In Proceedings of the 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, UK, 26–28 October 2015; pp. 2359–2364.
54. Basiri, M.E.; Kabiri, A. Sentence-level sentiment analysis in Persian. In Proceedings of the 2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA), Shahrekord, Iran, 19–20 April 2017; pp. 84–89.
55. Ebrahimi Rashed, F.; Abdolvand, N. A Supervised Method for Constructing Sentiment Lexicon in Persian Language. *J. Comput. Robot.* **2017**, *10*, 11–19.
56. Saraee, M.; Bagheri, A. Feature selection methods in Persian sentiment analysis. In *International Conference on Application of Natural Language to Information Systems*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 303–308.
57. Amiri, F.; Scerri, S.; Khodashahi, M. Lexicon-based sentiment analysis for Persian Text. In Proceedings of the International Conference Recent Advances in Natural Language Processing, Hissar, Bulgaria, 7–9 September 2015; pp. 9–16.
58. Vaziripour, E.; Giraud-Carrier, C.G.; Zappala, D. Analyzing the Political Sentiment of Tweets in Farsi. *ICWSM* **2016**, *10*, 699–702.
59. Sabeti, B.; Hosseini, P.; Ghassem-Sani, G.; Mirroshandel, S.A. LexiPers: An ontology based sentiment lexicon for Persian. *GCAI* **2016**, *41*, 329–339. [[CrossRef](#)]

60. Basiri, M.E.; Naghsh-Nilchi, A.R.; Ghassem-Aghaee, N. A framework for sentiment analysis in persian. *Open Trans. Inf. Process.* **2014**, *1*, 1–14. [[CrossRef](#)]
61. Alimardani, S.; Aghaie, A. Opinion Mining in Persian Language Using Supervised Algorithms. *J. Inf. Syst. Telecommun. (JIST)* **2015**, *11*. [[CrossRef](#)]
62. Day, M.Y.; Lin, Y.D. Deep Learning for Sentiment Analysis on Google Play Consumer Review. In Proceedings of the 2017 IEEE International Conference on Information Reuse and Integration (IRI), San Diego, CA, USA, 4–6 August 2017; pp. 382–388.
63. Paredes-Valverde, M.A.; Colomo-Palacios, R.; Salas-Zárate, M.d.P.; Valencia-García, R. Sentiment analysis in Spanish for improvement of products and services: A deep learning approach. *Sci. Program.* **2017**, *2017*, 1329281. [[CrossRef](#)]
64. Baniata, L.H.; Park, S.B. Sentence Representation Network for Arabic Sentiment Analysis. *Korean Tech. J.* **2016**, 470–472.
65. Xiao, K.; Zhang, Z.; Wu, J. Chinese text sentiment analysis based on improved Convolutional Neural Networks. In Proceedings of the 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 26–28 August 2016; pp. 922–926.
66. Dahou, A.; Xiong, S.; Zhou, J.; Haddoud, M.H.; Duan, P. Word embeddings and convolutional neural network for arabic sentiment classification. In Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers (COLING 2016), Osaka, Japan, 11–16 December 2016; pp. 2418–2427.
67. Le, T.A.; Moeljadi, D.; Miura, Y.; Ohkuma, T. Sentiment Analysis for Low Resource Languages: A Study on Informal Indonesian Tweets. In Proceedings of the 12th Workshop on Asian Language Resources (ALR12), Osaka, Japan, 12 December 2016; pp. 123–131.
68. Shuang, K.; Ren, X.; Chen, J.; Shan, X.; Xu, P. Combining Word Order and CNN-LSTM for Sentence Sentiment Classification. In *Proceedings of the 2017 International Conference on Software and e-Business*; ACM: New York, NY, USA, 2017; pp. 17–21.
69. Wehrmann, J.; Becker, W.; Cagnini, H.E.; Barros, R.C. A character-based convolutional neural network for language-agnostic Twitter sentiment analysis. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 2384–2391.
70. Sankar, H.; Subramaniaswamy, V.; Vijayakumar, V.; Arun Kumar, S.; Logesh, R.; Umamakeswari, A. Intelligent sentiment analysis approach using edge computing-based deep learning technique. *Softw. Pract. Exp.* **2019**, *50*, 645–657. [[CrossRef](#)]
71. Ali, N.M.; El Hamid, A.; Mostafa, M.; Youssif, A. Sentiment Analysis for Movies Reviews Dataset Using Deep Learning Models. *Int. J. Data Min. Knowl. Manag. Process. (IJDKP)* **2019**, *9*. Available online: <https://airconline.com/ijdkp/V9N3/9319ijdkp02.pdf> (accessed on 15 March 2021).
72. Lovins, J.B. Development of a stemming algorithm. *Mech. Transl. Comp. Linguist.* **1968**, *11*, 22–31.
73. Hussain, A.; Tahir, A.; Hussain, Z.; Sheikh, Z.; Gogate, M.; Dashtipour, K.; Ali, A.; Sheikh, A. Artificial intelligence-enabled analysis of UK and US public attitudes on Facebook and Twitter towards COVID-19 vaccinations. *medRxiv* **2020**. [[CrossRef](#)]
74. Gogate, M.; Dashtipour, K.; Hussain, A. Visual Speech In Real Noisy Environments (VISION): A Novel Benchmark Dataset and Deep Learning-based Baseline System. In Proceedings of the Interspeech 2020, Shanghai, China, 25–29 October 2020; pp. 4521–4525.
75. Guellil, I.; Adeel, A.; Azouaou, F.; Benali, F.; Hachani, A.E.; Dashtipour, K.; Gogate, M.; Ieracitano, C.; Kashani, R.; Hussain, A. A Semi-supervised Approach for Sentiment Analysis of Arab (ic+ izi) Messages: Application to the Algerian Dialect. *SN Comput. Sci.* **2021**, *2*, 1–18. [[CrossRef](#)]
76. Ahmed, R.; Dashtipour, K.; Gogate, M.; Raza, A.; Zhang, R.; Huang, K.; Hawalah, A.; Adeel, A.; Hussain, A. Offline Arabic Handwriting Recognition Using Deep Machine Learning: A Review of Recent Advances. In *International Conference on Brain Inspired Cognitive Systems*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 457–468.
77. Celikyilmaz, A.; Hakkani-Tür, D.; Feng, J. Probabilistic model-based sentiment analysis of twitter messages. In Proceedings of the 2010 IEEE Spoken Language Technology Workshop, Berkeley, CA, USA, 12–15 December 2010; pp. 79–84.
78. Voutilainen, A. Part-of-speech tagging. In *The Oxford Handbook of Computational Linguistics*; 2003. pp. 219–232. Available online: [https://books.google.com.sg/books?hl=zh-TW&lr=&id=y16AnaKtVAKC&oi=fnd&pg=PA219&dq=Voutilainen,+A.+Part-of-speech+tagging.+In+The+Oxford+Handbook+of+Computational+Linguistics&ots=\\_VTd79BLGn&sig=5\\_1oMxWkr3o-WimMcCWWC88cDNw#v=onepage&q=Voutilainen%2C%20A.%20Part-of-speech%20tagging.%20In%20The%20Oxford%20Handbook%20of%20Computational%20Linguistics&f=false](https://books.google.com.sg/books?hl=zh-TW&lr=&id=y16AnaKtVAKC&oi=fnd&pg=PA219&dq=Voutilainen,+A.+Part-of-speech+tagging.+In+The+Oxford+Handbook+of+Computational+Linguistics&ots=_VTd79BLGn&sig=5_1oMxWkr3o-WimMcCWWC88cDNw#v=onepage&q=Voutilainen%2C%20A.%20Part-of-speech%20tagging.%20In%20The%20Oxford%20Handbook%20of%20Computational%20Linguistics&f=false) (accessed on 15 March 2021).
79. Poria, S.; Chaturvedi, I.; Cambria, E.; Hussain, A. Convolutional MKL based multimodal emotion recognition and sentiment analysis. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 12–15 December 2016; pp. 439–448.
80. Gogate, M.; Dashtipour, K.; Adeel, A.; Hussain, A. CochleaNet: A robust language-independent audio-visual model for real-time speech enhancement. *Inf. Fusion* **2020**, *63*, 273–285. [[CrossRef](#)]
81. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. 1999. Available online: [https://digital-library.theiet.org/content/conferences/10.1049/cp\\_19991218](https://digital-library.theiet.org/content/conferences/10.1049/cp_19991218) (accessed on 15 March 2021).
82. Adeel, A.; Ahmad, J.; Hussain, A. Real-Time Lightweight Chaotic Encryption for 5G IoT Enabled Lip-Reading Driven Secure Hearing-Aid. *arXiv* **2018**, arXiv:1809.04966.
83. Gogate, M.; Adeel, A.; Marxer, R.; Barker, J.; Hussain, A. DNN driven speaker independent audio-visual mask estimation for speech separation. *arXiv* **2018**, arXiv:1808.00060.

84. Gogate, M.; Dashtipour, K.; Bell, P.; Hussain, A. Deep Neural Network Driven Binaural Audio Visual Speech Separation. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–7.
85. Ieracitano, C.; Adeel, A.; Gogate, M.; Dashtipour, K.; Morabito, F.C.; Larijani, H.; Raza, A.; Hussain, A. Statistical analysis driven optimized deep learning system for intrusion detection. In *International Conference on Brain Inspired Cognitive Systems*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 759–769.
86. Ghasemi, R.; Ashrafi Asli, S.A.; Momtazi, S. Deep Persian sentiment analysis: Cross-lingual training for low-resource languages. *J. Inf. Sci.* **2020**. [[CrossRef](#)]
87. Asli, S.A.A.; Sabeti, B.; Majdabadi, Z.; Golazizian, P.; Fahmi, R.; Momenzadeh, O. Optimizing annotation effort using active learning strategies: A sentiment analysis case study in persian. In Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; pp. 2855–2861.
88. Wang, W.; Wu, J. Notice of Retraction: Emotion recognition based on CSO&SVM in e-learning. In Proceedings of the 2011 Seventh International Conference on Natural Computation, Shanghai, China, 26–28 July 2011; Volume 1, pp. 566–570.