

Article

Multi-Aspect Oriented Sentiment Classification: Prior Knowledge Topic Modelling and Ensemble Learning Classifier Approach

Najwa AlGhamdi ^{1,*} , Shaheen Khatoon ² and Majed Alshamari ¹¹ Department of Information Systems, King Faisal University, Al-Ahsa 31982, Saudi Arabia; smajed@kfu.edu.sa² School of AI and Advanced Computing, Xi'an Jiaotong-Liverpool University, Suzhou 215000, China; Shaheen.Khatoon@xjtlu.edu.cn

* Correspondence: Najwa.h.alghamdi@gmail.com

Abstract: User-generated content on numerous sites is indicative of users' sentiment towards many issues, from daily food intake to using new products. Amid the active usage of social networks and micro-blogs, notably during the COVID-19 pandemic, we may glean insights into any product or service through users' feedback and opinions. Thus, it is often difficult and time consuming to go through all the reviews and analyse them in order to recognize the notion of the overall goodness or badness of the reviews before making any decision. To overcome this challenge, sentiment analysis has been used as an effective rapid way to automatically gauge consumers' opinions. Large reviews will possibly encompass both positive and negative opinions on different features of a product/service in the same review. Therefore, this paper proposes an aspect-oriented sentiment classification using a combination of the prior knowledge topic model algorithm (SA-LDA), automatic labelling (SentiWordNet) and ensemble method (Stacking). The framework is evaluated using the dataset from different domains. The results have shown that the proposed SA-LDA outperformed the standard LDA. In addition, the suggested ensemble learning classifier has increased the accuracy of the classifier by more than ~3% when it is compared to baseline classification algorithms. The study concluded that the proposed approach is equally adaptable across multi-domain applications.

Keywords: sentiment classification; prior knowledge; topic models; data labelling; ensemble learning; stacked generalization



Citation: AlGhamdi, N.; Khatoon, S.; Alshamari, M. Multi-Aspect Oriented Sentiment Classification: Prior Knowledge Topic Modelling and Ensemble Learning Classifier Approach. *Appl. Sci.* **2022**, *12*, 4066. <https://doi.org/10.3390/app12084066>

Academic Editors: Kristina Yordanova and Emma Tonkin

Received: 14 March 2022

Accepted: 12 April 2022

Published: 18 April 2022

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



Copyright: © 2022 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/>).

1. Introduction

Amid the active usage of social networks and micro-blogs, especially during the COVID-19 pandemic, we may glean insights into any product or service through users' feedback and opinions. Platforms such as micro-blogs, social media sites, online reviews, and discussion forums are rapidly growing. Therefore, it is challenging and time-consuming to go through all the reviews and analyse them with the intention of discovering the notion of the overall goodness or badness of these reviews. Accordingly, the essential endeavours to automatically analyse the sentiments of the users' reviews are increasingly needed.

Opinion mining and sentiment analysis are automatic classifications of textual information that focus on classifying data according to polarity (positive or negative). These automatic techniques could possibly be among the adopted ways to gauge both user impressions and satisfaction. User-generated content usually contains unstructured text that is used in classification tasks such as information extraction (IE), text analysis and natural language processing (NLP). It is applied to a vast number of reviews. Therefore, there is urgent demand for an advanced framework and formulas that can deal with the massive amount of information in order to precisely handle them and provide the most accurate related results.

However, predicting overall polarities for each review is not enough since the review could provide comments on various aspects of the corresponding product or service. For instance, one review about a restaurant may mention the prices, cleanliness, services and more. Analysing these aspects, rather than the overall review, constructs a better understanding of the exact leading pros and cons of the product or service. Therefore, the study focuses on performing aspect-level sentiment classification that predicts every aspect.

This paper proposes a multi-aspect-oriented sentiment classification model by using a combination of the prior knowledge topic model algorithm (SA-LDA), automatic labelling (SentiWordNet) and ensemble method (Stacking). In this study, the multi-aspect sentiment analysis is addressed by using a topic model and an ensemble learning method. However, the challenge of the models is that documents are rich in excessively informal and colloquial language. Thus, this research aims to identify an approach that depends on the combination of probabilistic topic modelling, namely, Seeded Aspect Latent Dirichlet Allocation (SA-LDA) and an ensemble learning method, to analyse and visualize noticeable aspects of text documents and classify them afterwards. Different domains, methods and classifiers have been used to address the aspect extraction and sentiment analysis tasks.

Furthermore, to evaluate the effectiveness of the proposed model, we conduct extensive experiments on three different domains of online reviews (movie, restaurant, and domestic Saudi Airlines reviews). The proposed model shows promising results. As far as we know, no previous research has proposed a model similar to our proposed one, which consists of three main modules: (1) LDA-based topic modelling; (2) sentiment lexicon (SentiWordNet); (3) the ensemble classifier (stacked generalization method).

Section 2 of this paper describes several related works, whereas the description of data collection, multi-aspect extraction model, proposed methodology, and ensemble learning algorithm are presented in Section 3. The findings and the conclusion with future works are presented in Sections 4 and 5, respectively.

2. Related Work

In this section, we offer a brief summary of the previous work in the context of aspect extraction via prior knowledge topic modelling, sentiment lexicon classification, and ensemble learning methods for Sentiment Analysis.

2.1. Multi-Aspect Topic Modelling for Aspect Extraction (Prior Knowledge Models)

Aspect extraction is one of the central phases in analysing the expressed opinions, emotions and viewpoints in textual data shared for a certain topic. Despite the current aspect extraction procedures that are based on topic models, the result of engaging only topic-models leads to generate unrelated and incoherent aspects. Prior knowledge semi-supervised models are introduced to enhance the correctness of aspects extraction using topic models with minimal user involvement. These proposed models aim to use domain-specific knowledge to guide the model in the topics extraction task to border the amount of unrelated extracted topics.

Several studies revealed that employing prior knowledge of a topic model has raised the aspect extraction accuracy. However, existing research studies have concentrated on a single domain using knowledge to extract aspects from a specific domain. For instance, Shi et al. [1] proposed a novel clustering method by leveraging prior knowledge to enhance the web services clustering task accuracy using a semi-supervised technique. The results have confirmed that the approach provides a major improvement in the clustering accuracy.

There is a considerable amount of literature on the prior knowledge topic model, especially with the LDA model, for instance, Concept-LDA [2], MC-LDA [3], SLM [4], MDK-LDA [5], GK-LDA [6], AKL [7], LTM [8], UFL-LDA [9], and many more.

The overall performance of all these and most other prior knowledge topic modelling techniques have used LDA-based techniques for aspect extraction to indicate that the extracted aspects are more corresponding and more accurate, as they significantly optimize the execution of the baseline topic models [10,11].

2.2. Sentiment Lexicon Classification

Sentiment lexicon classification (sentiment analysis) is the computational analysis of people's thoughts, ideas, and feelings towards an entity [12], and it involves classifying them into positive, neutral, or negative categories. Sentiment lexicon approaches are applied to label data and to measure the sentiment polarity. Sentiment lexicon classification relies on two sorts of approaches which are corpus-based and dictionary-based [13].

Many existing studies have applied sentiment lexicon to different domains and languages [14–18]. Most of these studies have used the lexicon SentiWordNet to extract sentiments and the results with little manual intervention. As it turns out, the chosen lexicon has improved the accuracy in terms of topic-specific lexical sentiments.

2.3. Ensemble Learning Method

Ensemble learning methods are among the top current research topics in machine learning [19]. Machine learning models are used for performing predictive classification in order to achieve a good performance, and special attention has been drawn to sentiment classification tasks. Some of the common ensemble learning methods include Averaging, Bagging, AdaBoost and Staking.

Many research studies investigated applying sentiment classification using ensemble methods [20–27]. Experiments were conducted on different domains such as restaurants [27–29], movies [30–33], products [34–37] and more. Additionally, the proposed ensemble models [23,38–42], with various characteristics such as domains, languages, and datasets have indicated that utilizing ensemble methods led to achieving optimized performance in the tasks of sentiment classification.

3. Materials and Methods

An overview of the proposed methodology is shown in Figure 1. It consists of data pre-processing followed by three core modules: (1) aspect extraction using the prior knowledge topic model (SA-LDA) algorithm; (2) automatic labelling (SentiWordNet); (3) ensemble learning classifier (Stacking). The details of each component are described in the following subsections.

3.1. Dataset and Pre-Processing

The first module of the proposed methodology consists of data collection and pre-processing. In this module, the data about users' opinions towards different aspects is collected from different online reviews on several domains. Table 1 shows the basic descriptive information of the three datasets used in the experimental analysis.

Table 1. Summary of the datasets.

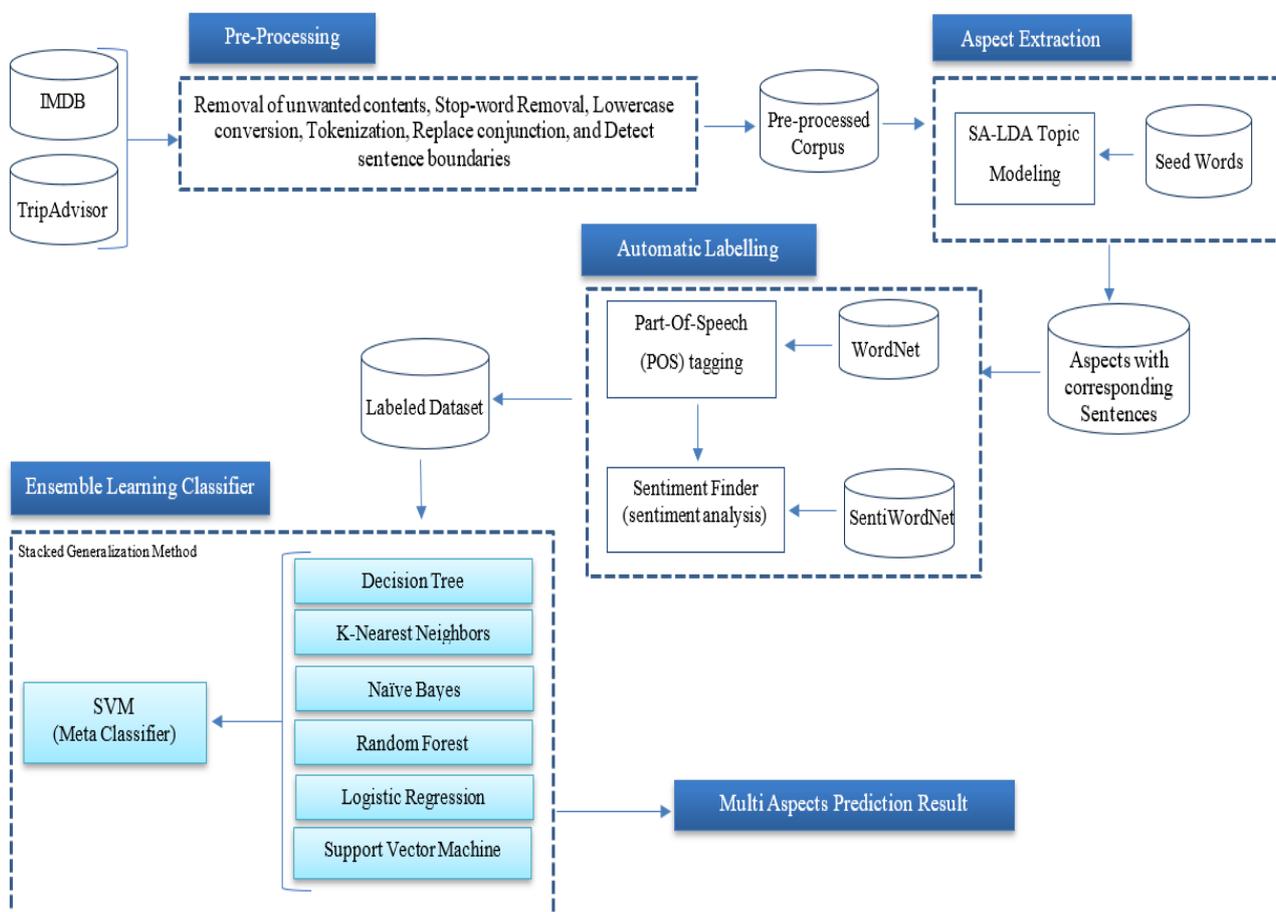
Datasets	No. of Reviews	Source of Datasets
Movie Reviews	2000	IMDB
Restaurant Reviews	2000	TripAdvisor
Domestic Saudi Airlines Reviews	2000	TripAdvisor
Total	6000 reviews	

A step-by-step procedure for data collection and pre-processing is outlined in Algorithm 1. The results were generated in a pre-processed textual corpus which contained an opinion unit (sentence) that would be ready to be handled to extract aspects and opinion aspects in the next step.

Algorithm 1: Algorithm for data collection and pre-processing**Input:** Online reviews (R_i)**Output:** Cleaned reviews (CR_i)For each Review in R_i , where $i = 1, 2, 3, 4 \dots n$

Apply:

1. Remove unwanted contents (R_i).
2. Remove Stop-word (R_i).
3. Converted to lowercase (R_i).
4. Tokenization (R_i).
5. Replace conjunction words with full-stop (R_i).
6. Detect sentence boundaries (R_i).
7. Repeat steps 2 to 5 until $R_i = R_n$.

**Figure 1.** The architecture of the proposed framework.

3.2. Aspect Extraction

The next step of the proposed model pipeline is automatically extracting semantic aspects (which are also called topics) from the pre-processed textual corpus. In this paper, a modified LDA model, called Seeded-Aspects LDA (SA-LDA), is proposed. It has an unlabelled pre-processed textual corpus that contains opinion units of a specific domain and an aspect specification as an input. An aspect specification is known as predefined aspects (seed words). In basic LDA, the model tends to only detect the most obvious aspects of a text corpus which may not cover the expected and desired aspects. Thus, we proposed a modified LDA model by providing seed words (seed aspects) to guide the model to only generate words from analogous seed aspects as presented in Figure 2.

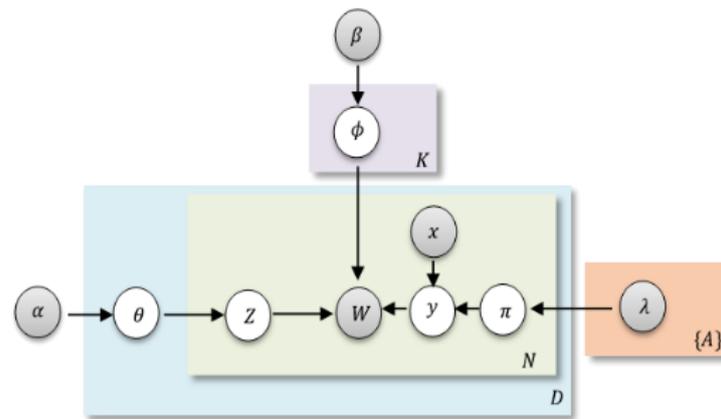


Figure 2. The proposed model in plate notation.

The SA-LDA at its basis comprises an LDA-based topic modelling, and it is extended with biased topic modelling hyper-parameters (β and α) that are based on continuous word embeddings. The number of aspects (k) is set based on the number of unique main aspects needed. Each review is modelled by an aspect and contains a sentence. The proposed model in plate notation is illustrated in Figure 2, where the generative hypothesis algorithm is described in Algorithm 2.

Algorithm 2: Algorithm for the generative hypothesis

- 1 For each aspect $k = 1 \dots K$,
 - Choose seed aspect $\varnothing_k^A \sim Dir(\beta)$.
- 2 For each review d ,
 - Choose $\theta_d \sim Dir(\alpha)$.
 - For each token $w_n, n = 1 \dots N_d$,
 - Sample $\pi_{d,n} \sim MaxEnt(\lambda, x_{w_{d,n}})$.
 - Draw a topic $Z_n \sim Multi(\theta_d)$.
 - Draw an indicator $y_{d,n} \sim Bern(\pi_{d,n})$
 - if $y_{d,n} = A$:
 - Sample a word $w_n \sim Multi(\varnothing_{Z_n}^A)$.

We provided the model with several seed words for each main aspect as shown in Table 2. After feeding in unique aspects and seeded words for each dataset, each review sentence becomes ready for the next phase of the sentiment analysis task as described in the next subsection.

Table 2. Aspects and seed words for each domain.

Domain	Aspect	Seed Words
Movie	ACT	performance, actor, play, character, role, scene
	PLOT	story, script, sequence, scenario
	SOUNDTRACK	sound, audio, music, playlist, effect
Restaurant	FOOD	taste, dish, dinner, appetizer, menu
	SERVICE	internet, parking, delivery, location, seating, staff
	PRICE	payment, price, discount, cost, pay offer
Domestic Saudi Airlines	FLIGHT	meal, internet, entertainment, seat, cleanliness, drink
	SERVICE	lounge, ticket, baggage, upgrade, punctuality
	STAFF	crew, captain, pilot, service, steward

3.3. Automatic Labelling System

Automatic labelling uses the sentiment lexicon approach to label data and to measure the sentiment polarity. In order to label a dataset in this work, SentiWordNet is applied. SentiWordNet is obtained from the WordNet dictionary where each word is associated with a numerical score. In this phase, for each sentence, the SentiWordNet dictionary is applied to determine the polarity of each word, and then the polarity of the whole sentence is calculated by adding the polarity of each word. If the word is not in the SentiWordNet dictionary, it is searched for in the WordNet dictionary. WordNet is an English language dictionary that contains synonym words gathered into a set called syn-set. Thus, the analogous words related to the word in WordNet are fetched and searched in the SentiWordNet dictionary such that their sentiment score is selected for polarity calculation. This procedure increases the efficiency and effectiveness of automatic labelling.

Furthermore, some words, called negation words, may affect the sentiment orientation of other words in the sentence. Negation words are those words that reverse the polarity of the sentence when occurring in it. For example, in the text “the food is not good”, the negation word “not” reverses the polarity of the sentence. To handle this issue, a negation is considered in the polarity calculation. The algorithm of the automatic labelling phase is illustrated in Algorithm 3.

Algorithm 3: Algorithm of the automatic labelling

Input: Sentences, SentiWordNet, WordNet, NegationWords

Output: Labelled Dataset

for each sentence S :

 taggedSentence = POS(S)

for each WordCandidate (verb, adverb, and adjective) in taggedSentence

 LookupSentiWordNet (WordCandidate)

if WordCandidate not in SentiWordNet

 LookupWordNet (WordCandidate)

else if WordCandidate > 0

 polarity (WordCandidate) \leftarrow positive

else if WordCandidate < 0

 polarity (WordCandidate) \leftarrow negative

else if

 polarity (WordCandidate) \leftarrow neutral

else (there is NegationWords near WordCandidate)

 polarity (WordCandidate) \leftarrow opposite (polarity (WordCandidate))

 PolarityScore += LookupSentiWordNet (WordCandidate)

 TotalWordCandidateCount++

 AveragePolarity = PolarityScore / TotalWordCandidateCount

if AveragePolarity > 0

 return 1

else

 return 0

The result demonstrates the label (1 for positive and 0 for negative) and the sentiment polarity. Then, it is used for the next phase, which is the ensemble learning classifier. The labelled dataset is used to train the classification model. The ensemble learning classifier method is used for sentiment classification. Precisely, in the ensemble method, stacked generalization is employed on different classifier algorithms as explained in the next sub-section.

3.4. Predicting Polarity of Largescale Social Data Using Supervised Learning (The Ensemble Learning Classifier Method)

An ensemble algorithm is trained on the labelled dataset to classify the unseen reviews as positive or negative on the go. Up-to-date numerous ensemble learning methods have been developed and introduced to enhance the performance of classification tasks. The major purpose of the ensemble models is to combine a set of classifiers with the intention of

achieving a better and more reliable predictive performance than a single classifier [43]. The focus will be on the capability of an ensemble model to generate a better result compared to each baseline classifier. In this experiment, a stacked generalization method was used, as shown in Figure 3, because it minimizes generalization error.

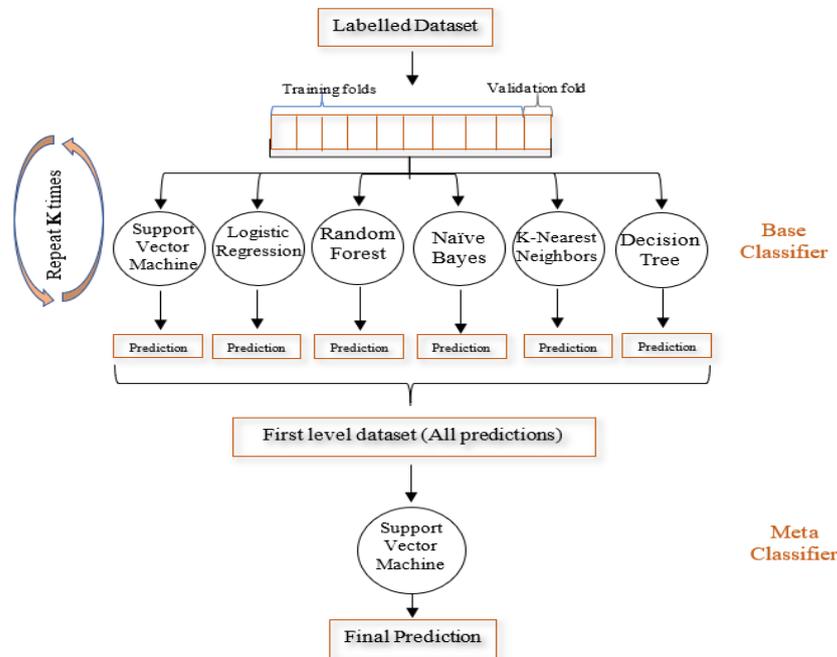


Figure 3. Steps of the ensemble learning.

The idea of stacked generalization is meant to combine the prediction result of several base classifiers in the first level using a meta classifier in the next level in order to minimize the generalization error. The process of performing a stacked generalization with k -fold cross-validation is shown in Figure 3.

The first step includes training the base classifiers in the first level, which are support vector machine, logistic regression, random forest, decision tree, naïve Bayes, and K-nearest neighbours by employing k -fold cross-validation on each classifier. The dataset is divided into k subsets. For each time in k sequential rounds, one of the k subsets is used as the test set and the other $k - 1$ subset is drawn from the training set. After that, each base classifier generates a prediction. Then, the prediction values from each classifier are combined and provided as the dataset for the second level. Finally, this step includes a training meta classifier on the second level with the first level dataset to produce the final prediction. Algorithm 4 describes the stacked generalization with k -fold cross-validation with $k = 10$.

Algorithm 4: Stacked Generalization with k -fold cross-validation

Input: Dataset D , Base classifiers t , base classifier prediction p , meta classifier m

Output: Ensemble Classifier Prediction P

Apply k -fold CV, $k = 10$, $D_n = \{D_1, D_2, \dots, D_{10}\}$ //Split the dataset into 10 subsets

for $k \leftarrow 1$ to n **do**

for each $t \leftarrow 1$ to T //base classifiers

 train the classifier p_{kt} from D_n .

end for

for D_p **do** //generate first level dataset

 get a dataset D_p , where $D_p = \{p_{t_1}, p_{t_2}, \dots, p_T\}$.

end for

train m from D_p //meta classifier

return P //final prediction

4. Evaluation Criteria and Experimental Results

The evaluation methods for classification models used in this paper are precision, recall and F-measure, as in [44]. They were used to estimate the performance result of each classifier. We evaluated our classifiers and models according to a 10-fold cross-validation scheme on the datasets.

In this section, we will evaluate and discuss the three main modules of the proposed model. In the first module (aspect extraction), we evaluated the proposed model, named SA-LDA topic modelling. This evaluation relies on two parts: (1) manual evaluation of each extracted aspect; (2) comparison of results with the based topic modelling algorithm regarding each domain.

In the second module (automatic labelling), we tested the accuracy of the proposed lexicon-based approach and verified the results with the manually labelled dataset. We also compared three lexicon-based approaches with the related works and the present results.

In the third module (ensemble classifier), we illustrated the performance of the proposed classifier model for the purpose of aspect sentiment analysis. This evaluation relies on two parts: (1) evaluating the performance and accuracy of the proposed model on three different domains; (2) comparing the proposed model to the baseline classifiers as well as another ensemble method.

4.1. Aspect Extraction (SA-LDA Model)

The result shows that SA-LDA extracts valuable aspects and relates them to the main aspect. However, LDA extracts many unrelated aspects along with some adjective words which are considered as opinion words more than aspects. Table 3 compares the results obtained from both models for each domain. The coloured words in 'red' indicate the errors or unrelated aspects. We manually evaluated the model based on the number of words that are related to the seed words/aspect which is our manual evaluation of the models. Even with these upsetting words, the proposed models can produce better results. However, the proposed model is flexible in a way that enables it to be adapted in any domain by specifying the seed words for the needed aspects.

Additionally, when the two results are compared, it is obvious that the proposed model outperforms the baseline model. Tables 3 and 4 illustrate the results of the performance of the two models in light of the three domains. Concerning the accuracy of SA-LDA, as illustrated in Table 4, it is clear that the Restaurant has the highest score with 86.7% while the Movie comes second with a score of 83.3%. Yet, Domestic Saudi Airline has the lowest score of 80%. Conversely, the standard model (LDA) scored lower accuracy results with 54%, 41% and 32% for Movie, Restaurant and Domestic Saudi Airlines, respectively. In conclusion, these results indicate that the proposed model has been more successful in detecting more correlated aspects, and it is likely to yield improved results with better performance.

4.2. Automatic Labelling (SentiWordNet)

Sentiment classification is an indication of the task of sentiment analysis which is a sub-field of natural language processing. The lexicon approach is applied to extract the opinion of each aspect by using SentiWordNet, which determines whether the text content specifies a positive or a negative review. Opinion extraction and automatic labelling are carried out in three steps: (1) applying part-of-speech tagging to each sentence; (2) extraction of all the opinion words and detecting the polarity of each opinion word; (3) looking for a negation word that is close to any opinion word, and once it is found, the polarity is reversed.

Opinion words are usually represented in the adjective, adverb, and verb forms such as "like" or "really" which affect the final result. For instance, the sentences "I like pizza" and "I really like pizza" both contain positive opinions, but the second sentence is more positive. Opinion words can be identified after applying POS tagging for each sentence, and it is typically found near the aspect.

The accuracy of SentiWordNet performance was measured by applying SVM classifier and five-fold cross-validation. The overall results of the accuracy for each domain are

shown in Table 5. The results are compared with the related work where SentiWordNet and SVM classifier have been used for different sentiment analysis tasks.

The results indicate that the accuracy of ‘Restaurant’ scores has recorded the highest percentage with 69.4%, while ‘Movie’ comes second with 65%, and the lowest score is recorded by the ‘Domestic Saudi Airline’ with 63.2%. The percentage distribution of the sentiment polarity for each aspect of the three domains is presented in Figure 4.

Table 3. Comparison between the proposed topic modelling results and the based topic modelling algorithm.

Domain	Aspect	LDA Model	SA-LDA Model
Movie	Act	(actor, show, play, character, art, life, movie , appear, sound, only)	(performance, act, actor, character, play, actress, part, scene, role, do)
	Plot	(story, play, series, know, role , line, set, sound, voice, say)	(story, series, play, script, sequence, role , scenario, line, do , text)
	Soundtrack	(music, song, sound, great , play, musical, product, show, story, movie)	(music, sound, audio, song, show , effect, playlist, soundtrack, play, movie)
Restaurant	Food	(food, restaurant, menu, good , chicken, dishes, street, visit, dinner, service)	(restaurant, menu, food, taste, appetizer, dinner, dishes, view , cook, flavor)
	Service	(staff, manager, ask , service, food, said, friendly, restaurant, told, eat)	(service, staff, order, internet, view, location, delivery, parking, menu , seating)
	Price	(price, card, feel, money , charged, payment, cheap, night, really, like)	(price, cost, payment, offer, card , discount, pay, bill, charge, worth)
Domestic Saudi Airlines	Flight	(ticket, check , flight, bad, lounge, schedule, time, only , food, hour)	(flight, seat, meal, internet, food, entertainment, movie, clean, drink, ticket)
	Service	(staff , desk, check, airport, out, arrival, hour, ready, late, wait)	(ticket, service, lounge, flight, baggage, schedule, staff , upgrade, offer, punctuality)
	Staff	(staff, friendly , service, direct, talk, helpful, front, desk, time, late)	(staff, service, facility , schedule, time, crew, ticket , pilot, flight, plane)

Table 4. The performance of the proposed model and standard model across three domains.

Domain	Accuracy (SentiWordNet)	Accuracy of Related Work
Movie	65%	53.33% [45], 79% [46]
Restaurant	69.4%	53% [47], 56% [48]
The Domestic Saudi Airlines	63.2%	65.2% [49], 46.41% [50]

Table 5. Performance evaluation of reviews using SVM and comparison of related work results.

Domain	Accuracy of the Proposed Model SA-LDA (LDA with Seed Words)	Accuracy of the Baseline Model (LDA without Seed Words)
Movie	83.3%	54%
Restaurant	86.7%	41%
The Domestic Saudi Airlines	80%	32%

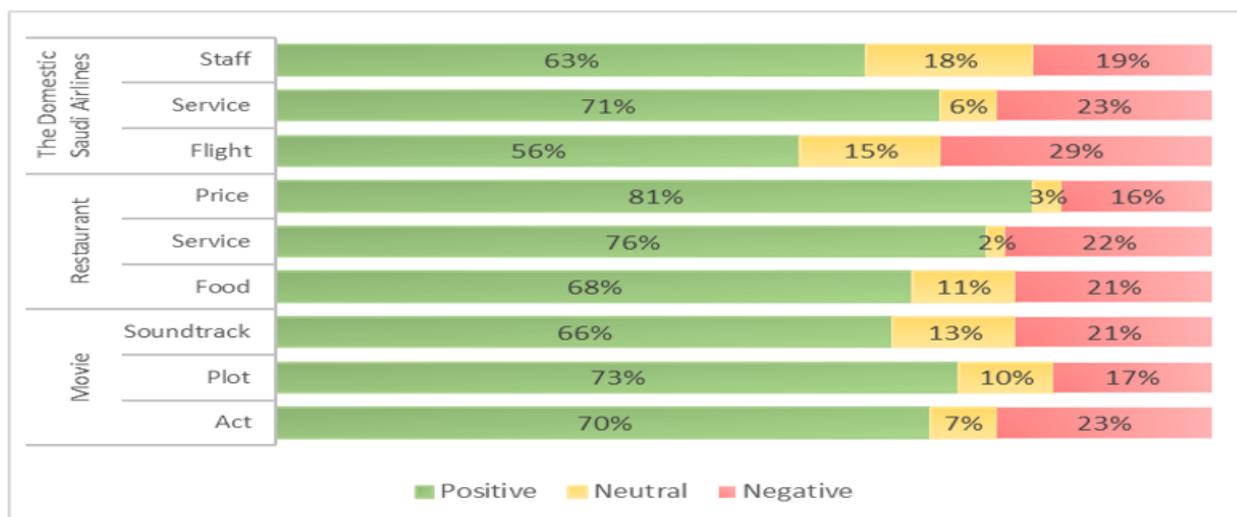


Figure 4. Distribution of sentiment polarity.

4.3. Ensemble Classifier (Stacking Generalization)

The performance evaluation of the proposed ensemble classifier model for the purpose of aspect sentiment analysis relies on two parts: (1) making a comparison between the proposed model and the baseline classifiers in addition to another ensemble method on three different domains; (2) evaluating the performance and accuracy of the proposed model on three different domains.

Tables 6 and 7 illustrate the comparison between the proposed model and the baseline classifiers as well as three other different ensemble methods including bagging, adaboost and majority voting for the selected domains.

As outlined in Tables 6 and 7, the proposed model has scored better results compared to the baseline classifiers and other ensemble classifier methods, with an accuracy level of 81.2%, precision of 81.1%, recall of 80.4%, and F1-scores of 81%. The lowest accuracy performance of other ensemble methods is for ‘majority voting’ with 77.5%. The lowest accuracy performance of the baseline classifier is ‘decision tree’ with 68.8%, whereas the highest accuracy result is 80.4% for the naïve Bayes classifier.

Table 6. Performance comparison of baseline classifiers on three various domains.

Domain	Base Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Restaurant	SVM	80.2	80.9	79.1	80
	LR	80.3	80.9	79	80
	RF	76	74.5	79	76.7
	DT	68.8	71.2	63.4	67.1
	NB	80.4	81	78.9	80.1
	KNN	68.9	65.9	78.5	71.6
Movie	SVM	81.2	81	80.7	80.3
	LR	80	71	73.5	72
	RF	73	73	72.6	72.7
	DT	72.2	69	75	71.8
	NB	71.5	80.5	64	71.3
	KNN	71	74	68.4	71
The Domestic Saudi Airlines	SVM	73.6	73	73	73.5
	LR	72	71	72.3	73
	RF	62	56	75	57.5
	DT	63.5	70.1	57	63
	NB	72	81	65	72
	KNN	73.2	76	70	73.4

Table 7. Performance comparison of different ensemble methods with proposed method on restaurant reviews.

Domain	Ensemble Method	Acc. (%)	P (%)	R (%)	F1 (%)
Restaurant	Bagging	80.3	80.1	79	80.1
	AdaBoost	79.5	81	76.4	78.8
	Majority Voting	77.5	76.4	79.4	77.9
	Stacked Generalization (Proposed)	83.2	83	82.4	83.1
Movie	Bagging	80	80.7	79.6	79.4
	AdaBoost	77.8	79	77.7	77.7
	Majority Voting	77.6	78	77.6	77.2
	Stacked Generalization (Proposed)	84	83.5	83	84
The Domestic Saudi Airlines	Bagging	74.1	74	74.3	74
	AdaBoost	77.5	77.3	76.2	76.3
	Majority Voting	75	74	75.3	75.7
	Stacked Generalization (Proposed)	84.4	83.1	82	84.7

5. Conclusions

The main aim of this paper is to develop an efficient model to discover sentiments associated with different aspects of a given text in order to make a more accurate decision from the users' perspective. The main objectives of the proposed system are: (1) Designing an efficient model to identify and extract all the possible aspects from given textual data. This is achieved by using natural language processing (NLP) to prepare the text in a format adopted by a topic model in addition to a topic model that extracts the main topics/aspects in that text. (2) Mapping between the extracted aspects and their opinions using linguistic and statistical techniques through utilizing a topic model and lexicon classification. (3) Developing a sentiment classification model in order to identify the sentiment orientation of the extracted aspect using an ensemble learning classifier.

To evaluate the performance of the proposed framework, we have compared each component to the baseline algorithms for the topic modelling, lexicon-based method and ensemble learning classifiers. The results have shown that the proposed framework is able to predict labels of the three review domains—restaurant, movie, and Saudi airlines—with an accuracy of 83.2%, 84% and 84.4% in each domain, respectively. Furthermore, once the proposed system is compared to the baselines algorithms, better results (higher than 2%) were scored in terms of the ability to predict the labels correctly.

This study has shown some promising results in the field of aspect-based sentiment analysis. It opened the windows wide for further research to enhance and expand this area of research. For future research, the proposed framework could be expanded to handle Arabic texts, which will be a challenging task. Likewise, future studies could apply more resources to the proposed framework to further enhance the results.

Author Contributions: Conceptualization, S.K. and M.A.; methodology, S.K. and N.A.; software, N.A.; validation, N.A.; formal analysis, N.A.; investigation, N.A. and S.K.; resources, N.A., M.A. and S.K.; data curation, N.A.; writing—original draft preparation, N.A.; writing—review and editing, S.K.; visualization, N.A.; supervision, S.K. and M.A.; funding acquisition, M.A.; project administration, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia for funding this research work through project number 523.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Shi, M.; Liu, J.; Cao, B.; Wen, Y.; Zhang, X. A Prior Knowledge Based Approach to Improving Accuracy of Web Services Clustering. In Proceedings of the 2018 IEEE International Conference on Services Computing (SCC), San Francisco, CA, USA, 2–7 July 2018.
2. Ekinci, E.; İlhan Omurca, S. Concept-LDA: Incorporating Babely into LDA for Aspect Extraction. *J. Inf. Sci.* **2020**, *46*, 406–418. [CrossRef]
3. Chen, Z.; Mukherjee, A.; Liu, B.; Hsu, M.; Castellanos, M.; Ghosh, R. Exploiting Domain Knowledge in Aspect Extraction. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; pp. 1655–1667.
4. Fang, L.; Huang, M. Fine Granular Aspect Analysis Using Latent Structural Models. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, Jeju, Korea, 8–14 July 2012; Volume 2, pp. 333–337.
5. Chen, Z.; Mukherjee, A.; Liu, B.; Hsu, M.; Castellanos, M.; Ghosh, R. Leveraging Multi-Domain Prior Knowledge in Topic Models. In Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, Beijing, China, 3–9 August 2013; pp. 2071–2077.
6. Chen, Z.; Mukherjee, A.; Liu, B.; Hsu, M.; Castellanos, M.; Ghosh, R. Discovering Coherent Topics Using General Knowledge. In Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management—CIKM '13, San Francisco, CA, USA, 27 October–1 November 2013; pp. 209–218.
7. Chen, Z.; Mukherjee, A.; Liu, B. Aspect Extraction with Automated Prior Knowledge Learning. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MD, USA, 22–27 June 2014; pp. 347–358.
8. Chen, Z.; Liu, B. Topic Modeling Using Topics from Many Domains, Lifelong Learning and Big Data. In Proceedings of the the 31st International Conference on Machine Learning, Beijing, China, 21–26 June 2014; Volume 32, pp. II-703–II-711.
9. Wang, T.; Cai, Y.; Leung, H.; Lau, R.Y.K.; Li, Q.; Min, H. Product Aspect Extraction Supervised with Online Domain Knowledge. *Knowl.-Based Syst.* **2014**, *71*, 86–100. [CrossRef]
10. Rana, T.A.; Cheah, Y.-N.; Letchmunan, S. Topic Modeling in Sentiment Analysis: A Systematic Review. *J. ICT Res. Appl.* **2016**, *10*, 76–93. [CrossRef]
11. Majumder, N.; Bhardwaj, R.; Poria, S.; Zadeh, A.; Gelbukh, A.; Hussain, A.; Morency, L.-P. Improving Aspect-Level Sentiment Analysis with Aspect Extraction. *Neural Comput. Appl.* **2020**, *2021*, 1–14. [CrossRef]
12. Medhat, W.; Hassan, A.; Korashy, H. Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Eng. J.* **2014**, *5*, 1093–1113. [CrossRef]
13. Khatoun, S.; Romman, L.A. Domain Independent Automatic Labeling System for Large-Scale Social Data Using Lexicon and Web-Based Augmentation. *ITC* **2020**, *49*, 36–54. [CrossRef]
14. Keshavarz, H.; Abadeh, M.S. ALGA: Adaptive Lexicon Learning Using Genetic Algorithm for Sentiment Analysis of Microblogs. *Knowl.-Based Syst.* **2017**, *122*, 1–16. [CrossRef]
15. Yang, L.; Li, Y.; Wang, J.; Sherratt, R.S. Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. *IEEE Access* **2020**, *8*, 23522–23530. [CrossRef]
16. Liapakis, A. A Sentiment Lexicon-Based Analysis for Food and Beverage Industry Reviews. The Greek Language Paradigm. *SSRN J.* **2020**, *9*, 21–42. [CrossRef]
17. Zhang, S.; Wei, Z.; Wang, Y.; Liao, T. Sentiment Analysis of Chinese Micro-Blog Text Based on Extended Sentiment Dictionary. *Future Gener. Comput. Syst.* **2018**, *81*, 395–403. [CrossRef]
18. Bandhakavi, A.; Wiratunga, N.; Padmanabhan, D.; Massie, S. Lexicon Based Feature Extraction for Emotion Text Classification. *Pattern Recognit. Lett.* **2017**, *93*, 133–142. [CrossRef]
19. Kuncheva, L.I. *Combining Pattern Classifiers: Methods and Algorithms*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014; ISBN 978-1-118-91456-4.
20. Onan, A.; Korukoğlu, S.; Bulut, H. A Multiobjective Weighted Voting Ensemble Classifier Based on Differential Evolution Algorithm for Text Sentiment Classification. *Expert Syst. Appl.* **2016**, *62*, 1–16. [CrossRef]
21. Oussous, A.; Lahcen, A.A.; Belfkih, S. Improving Sentiment Analysis of Moroccan Tweets Using Ensemble Learning. In *Big Data, Cloud and Applications*; Tabii, Y., Lazaar, M., Al Achhab, M., Enneya, N., Eds.; Communications in Computer and Information Science; Springer International Publishing: Cham, Switzerland, 2018; Volume 872, pp. 91–104, ISBN 978-3-319-96291-7.
22. Nehe, M.P.B.; Nawathe, A. Aspect Based Sentiment Classification Using Machine Learning for Online Reviews. 2020. Available online: [https://easychair.org/publications/preprint_download/xnVW](https://easychair.org/publications/preprint/download/xnVW) (accessed on 13 March 2022).
23. Shoukry, A.; Rafea, A. Machine Learning and Semantic Orientation Ensemble Methods for Egyptian Telecom Tweets Sentiment Analysis. *JWE* **2020**, *19*, 195–214. [CrossRef]
24. Sultana, N.; Islam, M.M. Meta Classifier-Based Ensemble Learning for Sentiment Classification. In *Proceedings of International Joint Conference on Computational Intelligence*; Uddin, M.S., Bansal, J.C., Eds.; Algorithms for Intelligent Systems; Springer: Singapore, 2020; pp. 73–84, ISBN 9789811375637.
25. Basiri, M.E.; Abdar, M.; Cifci, M.A.; Nemati, S.; Acharya, U.R. A Novel Method for Sentiment Classification of Drug Reviews Using Fusion of Deep and Machine Learning Techniques. *Knowl.-Based Syst.* **2020**, *198*, 105949. [CrossRef]
26. Khalid, M.; Ashraf, I.; Mehmood, A.; Ullah, S.; Ahmad, M.; Choi, G.S. GBSVM: Sentiment Classification from Unstructured Reviews Using Ensemble Classifier. *Appl. Sci.* **2020**, *10*, 2788. [CrossRef]

27. Tharwat, A. Classification Assessment Methods. *ACI* **2021**, *17*, 168–192. [[CrossRef](#)]
28. Raju, K.D.; Jayasingh, B.B. Machine Learning for Sentiment Analysis for Twitter Restaurant. *JES* **2018**, *9*, 21–27.
29. Waikul, V.; Ravgan, O.; Pavate, A. Restaurant Review Analysis and Classification Using SVM. *IOSR JEN* **2019**, *1*, 49–52.
30. Shariieff, H.; Sindhu, T.; SaiRamesh, L. Comparison of Machine Learning Techniques for Sentimental Analysis on Restaurant Reviews. *IJAEM* **2020**, *2*, 740–743.
31. Bandana, R. Sentiment Analysis of Movie Reviews Using Heterogeneous Features. In Proceedings of the 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), Kolkata, India, 4–5 May 2018; pp. 1–4.
32. Ghosh, M.; Sanyal, G. An Ensemble Approach to Stabilize the Features for Multi-Domain Sentiment Analysis Using Supervised Machine Learning. *J. Big Data* **2018**, *5*, 44. [[CrossRef](#)]
33. Untawale, T.M.; Choudhari, G. Implementation of Sentiment Classification of Movie Reviews by Supervised Machine Learning Approaches. In Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 27–29 March 2019; pp. 1197–1200.
34. Chang, J.-R.; Liang, H.-Y.; Chen, L.-S.; Chang, C.-W. Novel Feature Selection Approaches for Improving the Performance of Sentiment Classification. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *2021*, 1–14. [[CrossRef](#)]
35. Jagdale, R.S.; Shirsat, V.S.; Deshmukh, S.N. Sentiment Analysis on Product Reviews Using Machine Learning Techniques. In *Cognitive Informatics and Soft Computing*; Advances in Intelligent Systems and Computing Book Series; Springer: Berlin/Heidelberg, Germany, 2019; Volume 768, pp. 639–647.
36. Shaheen, M. Sentiment Analysis on Mobile Phone Reviews Using Supervised Learning Techniques. *IJMECS* **2019**, *11*, 32–43. [[CrossRef](#)]
37. Choudhari, P.; Veenadhari, S. Sentiment Classification of Online Mobile Reviews Using Combination of Word2vec and Bag-of-Centroids. In *Machine Learning and Information Processing*; Swain, D., Pattnaik, P.K., Gupta, P.K., Eds.; Advances in Intelligent Systems and Computing; Springer: Singapore, 2020; Volume 1101, pp. 69–80, ISBN 9789811518836.
38. Xu, F.; Pan, Z.; Xia, R. E-Commerce Product Review Sentiment Classification Based on a Naïve Bayes Continuous Learning Framework. *Inf. Process. Manag.* **2020**, *57*, 102–221. [[CrossRef](#)]
39. Al-Azani, S.; El-Alfy, E.-S.M. Using Word Embedding and Ensemble Learning for Highly Imbalanced Data Sentiment Analysis in Short Arabic Text. *Procedia Comput. Sci.* **2017**, *109*, 359–366. [[CrossRef](#)]
40. Khan, J.; Alam, A.; Hussain, J.; Lee, Y.-K. EnSWF: Effective Features Extraction and Selection in Conjunction with Ensemble Learning Methods for Document Sentiment Classification. *Appl. Intell.* **2019**, *49*, 3123–3145. [[CrossRef](#)]
41. Khai Tran; Thi Phan Deep Learning Application to Ensemble Learning—The Simple, but Effective, Approach to Sentiment Classifying. *Appl. Sci.* **2019**, *9*, 2760. [[CrossRef](#)]
42. İzmir Katip Çelebi Üniversitesi; Onan, A. Ensemble of Classifiers and Term Weighting Schemes for Sentiment Analysis in Turkish. *SRC* **2021**, *1*, 1–12. [[CrossRef](#)]
43. Ruta, D.; Gabrys, B. Classifier Selection for Majority Voting. *Inf. Fusion* **2005**, *6*, 63–81. [[CrossRef](#)]
44. Novaković, J.D.; Veljović, A.; Ilić, S.S.; Papić, Ž.; Milica, T. Evaluation of Classification Models in Machine Learning. *Theory Appl. Math. Comput. Sci.* **2017**, *7*, 39–46.
45. Bhoir, P.; Kolte, S. Sentiment Analysis of Movie Reviews Using Lexicon Approach. In Proceedings of the 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICIC), Madurai, India, 10–12 December 2015; pp. 1–6.
46. Rajeswari, A.M.; Mahalakshmi, M.; Nithyashree, R.; Nalini, G. Sentiment Analysis for Predicting Customer Reviews Using a Hybrid Approach. In Proceedings of the 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), Cochin, India, 2–4 July 2020; pp. 200–205.
47. Guha, S.; Joshi, A.; Varma, V. SIEL: Aspect Based Sentiment Analysis in Reviews. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, CO, USA, 4–5 June 2015; pp. 759–766.
48. Fikri, M.; Sarno, R. A Comparative Study of Sentiment Analysis Using SVM and SentiWordNet. *IJEECS* **2019**, *13*, 902–909. [[CrossRef](#)]
49. Yuan, P. Sentiment Classification and Opinion Mining on Airline Reviews. 2016. Available online: <https://www.semanticscholar.org/paper/Sentiment-Classification-and-Opinion-Mining-on-Yuan/daf1d9de4066eed1d193847cae578389da16c5e8> (accessed on 13 March 2022).
50. Mehta, P.; Chandra, S. Enhancement of SentiWordNet Using Contextual Valence Shifters. *IJDATS* **2019**, *11*, 337. [[CrossRef](#)]