

## Article

# Justifying Arabic Text Sentiment Analysis Using Explainable AI (XAI): LASIK Surgeries Case Study

Younna Abdelwahab <sup>\*</sup>, Mohamed Kholief  and Ahmed Ahmed Hesham Sedky

College of Computing, Arab Academy for Science, Technology, and Maritime Transport, Alexandria 1029, Egypt  
\* Correspondence: Youmna.Abdelfattah5@student.aast.edu

**Abstract:** With the increasing use of machine learning across various fields to address several aims and goals, the complexity of the ML and Deep Learning (DL) approaches used to provide solutions has also increased. In the last few years, Explainable AI (XAI) methods to further justify and interpret deep learning models have been introduced across several domains and fields. While most papers have applied XAI to English and other Latin-based languages, this paper aims to explain attention-based long short-term memory (LSTM) results across Arabic Sentiment Analysis (ASA), which is considered an uncharted area in previous research. With the use of Local Interpretable Model-agnostic Explanation (LIME), we intend to further justify and demonstrate how the LSTM leads to the prediction of sentiment polarity within ASA in domain-specific Arabic texts regarding medical insights on LASIK surgery across Twitter users. In our research, the LSTM reached an accuracy of 79.1% on the proposed data set. Throughout the representation of sentiments using LIME, it demonstrated accurate results regarding how specific words contributed to the overall sentiment polarity classification. Furthermore, we compared the word count with the probability weights given across the examples, in order to further validate the LIME results in the context of ASA.

**Keywords:** deep learning; LSTM; Arabic sentiment analysis; Explainable AI; text mining



**Citation:** Abdelwahab, Y.; Kholief, M.; Sedky, A.A.H. Justifying Arabic Text Sentiment Analysis Using Explainable AI (XAI): LASIK Surgeries Case Study. *Information* **2022**, *13*, 536. <https://doi.org/10.3390/info13110536>

Academic Editor: Gabriele Gianini

Received: 9 October 2022

Accepted: 8 November 2022

Published: 11 November 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

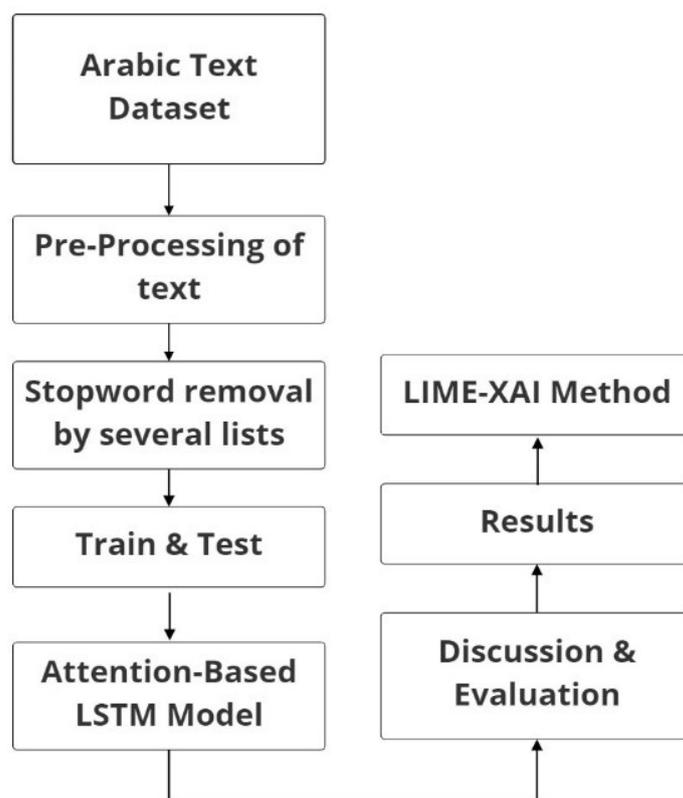
In the last couple of years, machine learning approaches have been applied successfully throughout a wide range of applications, such as medical diagnostics, hospitality, and other domain-specific fields. While the associated models have been improving over time, the complexity of each model has also continued to increase. Furthermore, despite these models increasing in popularity, many still lack explanation. As has been stated in [1], the main purpose of applying XAI is to answer one or more of the main seven goals, including reliability, usability, trust, fairness, privacy, causality, and transparency. Therefore, XAI has been used across different deep learning models in order to further justify the proposed classification within a specific domain's functionality, as well as the overall reliability of Deep Learning and Machine Learning [2]. As previously stated, machine learning has been applied for various purposes. One such application—sentiment analysis—involves determining the polarity of a text as negative, neutral, or positive [3].

Throughout previous research, sentiment analysis has been applied through the use of ML and DL models for accurate polarity classification in different domains and languages. For example, in [4], the authors have proposed a sentiment analysis model to classify the polarity of customer reviews on a Chinese-based e-commerce website. They collected about 100,000 customer reviews to perform the testing and training. Meanwhile, in [5], the authors used sentiment analysis to measure the destination carrying capacity targeted at a specific city in Europe, using online reviews from TripAdvisor.

Even though most research papers have targeted the English language, some studies have considered Arabic text sentiment analysis (ASA) as well. In a previous review on ASA [6], it has been stated that ASA is challenging due to the different dialects and

morphology of the Arabic language, as well as the imitation faced along the process. Some studies have been implemented in Arabic; for example, in [7], the authors implemented sentiment analysis to assess various Twitter data regarding COVID-19. The authors of this particular study used the proposed model as a precautionary measure, rather than a measure of being a potential COVID-19 patient, which was adjusted to predict the individual perceptions of Arabian users. In all of these studies, the researcher's main goal was to create an advanced model for the purpose of accurately classifying the polarity of text across social media services. The proposed models do not provide a comprehensible justification of how classification into different polarities is carried out. Therefore, some studies have begun to apply XAI methods to further justify the DL model results.

Throughout our previous works, several experiments have been carried out on several DL models across different sentiment levels, which led us to conclude that the attention-based LSTM has the best performance across the Arabic data set in terms of word-level sentiment analysis [8]. Therefore, in this paper, we propose the application of an XAI method—LIME—to the attention-based LSTM model on an Arabic text data set concerning LASIK surgeries across Twitter users. The general approach used in this study is depicted in Figure 1.



**Figure 1.** Flowchart of Applied Approach.

The remainder of this paper is structured as follows: In Section 2, we provide the literature review, while Section 3 gives the background related to this study. Section 4 details the methodology used. Finally, Sections 5 and 6 are dedicated to the discussion and conclusion, respectively.

## 2. Literature Review

In previous studies on sentiment analysis, the use of DL models has been proposed. For example, in [7], a DL model was developed for COVID-19-related tweets; however, this study lacked an XAI model to further interpret the model classification process. Meanwhile, the authors in [9] have proposed an XAI-based NB model to better explain the results for COVID-19, by looking at the symptoms that were disclosed in Twitter tweets based in

Turkey in order to determine the approximate numbers of infected people and predict possible virus breakouts. In [10], the authors have provided a comparison of the XAI methods LIME and LRP through simulatability tests on English text, which led to the conclusion that both methods can help in increasing the understanding of the DL model.

Moreover, XAI has been used across different domains; for example, in [11], XAI methods on ML combined with LSTM have been used to predict stocks and further explain the sentiments of headlines that influence users, using LIME to enhance stock prediction. Meanwhile, in [12], the researchers have used LIME and SHAP (Shapley Additive explanations) values to validate the features used in order to defend a specific sentiment polarity obtained by LSTM and hybrid LSTM-based models on customer reviews of food services during the COVID-19 crisis. In another study, they aimed to further explain the sentiments of Twitter users by using the XAI method LIME on a proposed BI-LSTM model, in order to interpret public perception in several domains [13]. Within the NLP domain, such a model has been used to detect sarcasm, due to its complexity within the English text data set, by applying LIME and SHAP values on an ensemble-supervised learning algorithm, in order to elaborate how the model with selected features detects whether the text contains sarcasm [14]. Meanwhile, the authors in [15] have aimed to classify IT jobs using attention-based LSTM and finalized their work by comparing the word frequency outcome with the LIME prediction, which led to the conclusion that LIME helped in discovering a new way of identifying job descriptions.

Moreover, the authors in [16] have used LIME to clearly justify the classification of source code vulnerability detection by applying both ML and DL models on LIME. This led to the conclusion that LIME works well in vulnerability detection, with the limitation of not identifying the second IF condition in a code sample. In [17], XAI has been utilized for the classification of offensive text across topics in Bangla, which resulted in a graphical presentation of the topics that contained the majority of the offensive text. Finally, in [18], the authors have utilized an XAI method to help in understanding why a tweet text would be considered xenophobic/racist, in order to prevent xenophobic acts or events. Throughout the previous studies across several domains, LIME was utilized and presented accurate results when paired with LSTM and attention-based LSTM when a low-resource language was used [17], or for the detection of racism [18]. According to [6], a lot of challenges are faced in ASA due to variant dialects and slang used across social media services, which makes it hard to emphasize which word has the largest contribution when considering the polarity of each sentiment. While most papers have aimed to use the LIME approach for their proposed LSTM model concerning different domains within the English language, due to the enormous number of data sets and accessible corpus (see Table 1), fewer studies have applied it to lower-resource languages, specifically Arabic text. Given that the overall performance of LIME in previous studies was promising, in terms of justifying DL models, enhancing human interpretation of the model, and indicating its important features, in this study we apply LIME to an Arabic text sentiment analysis approach. To the best of our knowledge, this paper is one of the first works to apply an XAI approach to Arabic text. This paper aims to do so by applying it to a domain-specific data set regarding LASIK surgery, further justifying the sentiment analysis classification by applying the XAI model LIME to justify why certain features have been specified to a particular polarity. This paper contributes to the area by applying XAI to Twitter, carrying out Arabic sentiment analysis using an Arabic text data set, as well as through the creation of a data set on LASIK surgery in Arabic text across Twitter users.

**Table 1.** Previous work in which LIME was applied to DL Models.

Reference	Year	Scope	Classifiers	XAI Algorithm	Accuracy
In Hyeok Choi et al. [15]	2020	IT Job classification	LSTM, Attention-based LSTM	LIME	76%/91%
Aljameel et al. [9]	2021	Predict the possible outbreak of COVID-19 patients in turkey	NB	Probabilistic methods	93.6%
Gite et al. [11]	2021	Stock Prediction	ML and LSTM	LIME	NA
Chowdhury et al. [13]	2021	Interpret Sentiments across several domains of Twitter users	BI-LSTM	LIME	72%
Kumar et al. [14]	2021	Detecting Sarcasm	XGBoost	SHAP, LIME	NA
Tang, G. et al. [16]	2021	Source code vulnerability detection	LR, DT, SVM, and Bi-LSTM.	LIME	NA
Rathore et al. [10]	2022	Better classification of tweets in the English language	ANN	LIME, LRP	85%/90%
Adak A et al. [12]	2022	Validate features used to defend a specific sentiment polarity on food reviews	LSTM, Bi-LSTM, Bi-Gru-LSTM-CNN	SHAP LIME	96.7%, 95.85%, 96.33%
Aporna et al. [17]	2022	Classifying offensive speech in Bangla text	SVM, CNN, Bi-LSTM, Conv-LSTM	Graphical representation	67%/73%/75%/78%

### 3. Background

#### 3.1. XAI Tools

For the application of XAI across different fields and domains, the two most commonly used algorithms are LIME and SHAP. First, Local Interpretable Model-agnostic Explanation (LIME) is an open-source framework used to describe individual predictions of a machine learning model, first introduced in [19], which aimed to concentrate on the decision-making of complex ML algorithms and how humans can trust their predictions. Local means that the framework analyzes a specific observation. Interpretable means that the user should be able to understand the behavior of the model. Explanation indicates the output that the LIME framework produces. Meanwhile, Shapley Additive Explanation (SHAP) methods are used to describe how each feature affects the model and how it enables the global analysis of data sets, which is based on a game-theoretic approach, in order to explain the output of machine learning models [20]. ContrXT is a global proposed approach that traces the decision criteria of text classifiers by encoding changes in the decision logic and provides a global model agnostic Time contrastive explanation in natural language processing [21]. In this paper, they proposed a novel self-explaining architecture for neural network text classifiers based on both local and global interpretability in a single framework on sentences rather than words, which resulted in promising results [22]. In this paper, they proposed an approach to measure how correct the explanations provided by the local explanation method are in relation to the synthetic ground truth explanation. Experimental results demonstrate how the proposed approach can easily assess the local explanation of a site and characterize the quality of the local explanation method. Throughout their evaluation, this was tested across text, image and tabular data returning features and rules. The results of the local explanation of the word importance explanation on text stated that LIME extracts more stable explanation and resulted in higher recall and precision compared to SHAP. Moreover, the results returned the best explanation according to the words identified with respect to the number of words used as a vocabulary [23]. Additionally, the variations of dialects within the Arabic language that are used across the social media platforms will result in better explanation across the Arabic language.

Throughout previous works, we have observed that LIME works well with text data sets within the English language across different domains, which led to its usage in this research across Arabic text. LIME was chosen to further explain how the attention-based LSTM model classifies the polarity of ASA text, due to its nature as a Local Explainer, which is very helpful when using it across a language with a complex morphology and variant dialects. In the result, the representation can emphasize the importance of a word in a

single sentiment and explain how it was classified as a specific polarity, which is important since—in the case of Arabic words—a lot of variance in meaning may occur across dialects.

### 3.2. LASIK Surgeries

LASIK is a form of refractive surgery that can correct vision in people with near-sightedness, far-sightedness, or astigmatism. This is one of many vision correction surgeries that involves re-shaping the cornea—the clear area in front of the eye—so that light is focused on the retina (at the back of the eye) [24]. This surgery is popular among relevant patients across the world, particularly in Arabic-speaking countries. This has led to many questions regarding detailed information, recommendations, and sharing of previous experiences with LASIK surgery procedures, in order to further understand and prepare for the surgery, including asking optometrists about the variations in the LASIK Surgeries available for patients and which is more suitable.

## 4. Materials and Methods

### 4.1. Data Set Creation

For this research, we created a data set that was used throughout the experiments. Twitter provides different types of application programming interfaces (APIs). We used the Full Archive API, which is a premium service provided by Twitter. By using the Tweepy Python module, 10,000 Arabic tweets were successfully scraped for further processing and analysis. The scraping operation targeted the timeframe between January 2017 and December 2021. The number of records was narrowed down to 4201 after precise cleaning and initial pre-processing of retweets, unrelated tweets, and spam. The resulting remainder were records that consisted of text, written mainly in the Egyptian and Saudi dialects, MSA, and other dialects within Arabic-speaking countries. This was collected regarding a specific topic—LASIK Surgeries—using specific keywords. The first keyword was “ليزك”, and the second keyword was “تصحيح”. This particular topic was chosen due to its importance across the Middle East and the satisfactory nature of associated results across medical studies [25]. Furthermore, this provides a basis for the creation of an Arabic data set for common eye surgeries in the Middle East, which can be used across future studies, rather than Arabic-text data sets without a specific context. This data set is concentrated on Arabic-speaking Twitter users, and the data were labeled positive, negative, or neutral using a script and by manual curation for further accuracy. Furthermore, the data set is publicly available [26]. Table 2 shows the number of tweets per label in the created data set.

**Table 2.** Tweets per label in LASIK Surgery Data set.

	Positive	Negative	Neutral
Data set [21]	2355	1040	807

### 4.2. Data Pre-Processing

To fit our proposed approach, data pre-processing was conducted to clean the input tweet data. To simplify and standardize our text, we first removed all English and other Latin-based characters. Second, as URLs and links—which do not provide any necessary information—are commonly used to refer to any uniform resource or other Twitter users on the internet, they were removed using regular expressions (regex). Third, some common Arabic stop words, which do not contribute much information in the overall sentence, were removed. To filter and avoid these stop words, we used the NLTK package for Arabic text on the collected data set. Fourth, all punctuation was removed, except for the question and exclamation marks, due to their use in changing the overall meaning and conveying the message. Fifth, when dealing with texts, numbers may not add much information; as such, we eliminated them utilizing the re.sub module. Finally, repeated characters were not removed, due to their use in emphasizing or showing a particular feeling. For example, the



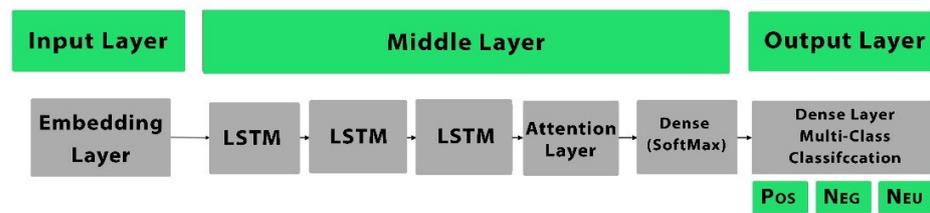


Figure 3. LSTM Model Framework.

4.5. Applying LIME XAI Method

XAI was applied to further explain and provide transparency for the applied sentiment analysis carried out on the data set. Similar approaches have been reported in previous works, such as in [8], where they used information about the symptoms written in Twitter posts to determine whether a user had potentially been exposed to the COVID-19 virus, in order to estimate places where a viral breakout could occur. To the best of our knowledge, this paper is the first of its kind to implement such an approach to ASA. Furthermore, the LIME (Local interpretable model-agnostic explanations) XAI model [28] was applied, due to its model-agnostic nature, which makes it suitable for use with various other models. This approach acts as an approximation technique for the DL model, using a local, interpretable model to explain each prediction. First, we applied it to the Arabic text data set that specifically targets the general opinion of users regarding LASIK surgeries with positive sentiment, in order to further understand some of the potential concerns and thoughts of users across social media. As previously mentioned, the data set consisted of posts with positive, negative, and neutral labels, for a total of 4202 texts. In the experiments previously carried out, the attention-based LSTM achieved higher accuracy. We applied the XAI method, which randomly sampled from the LSTM model to further explain why they were classified with a specific sentiment, considering the representation of sentiments that could be easy for non-native speakers to comprehend when translated and interpreted. Figure 4 shows an illustration of a sentence that was originally classified as having positive sentiment. The original sentence states that “his eyesight was weak and soon he will gain back his full eyesight.” Even though the LASIK surgery keyword was not mentioned within the sentence, it indicated the perceived recovery of his eyesight after performing the surgery. This is represented in the illustration below where the words “نظره” and “يشوف” were categorized as positive and, even though the words “يرجع”, “هيعمل”, and “ضعف” were classified as negative, within the Arabic morphology, they could have a double meaning, depending on other words within the text (as shown below).

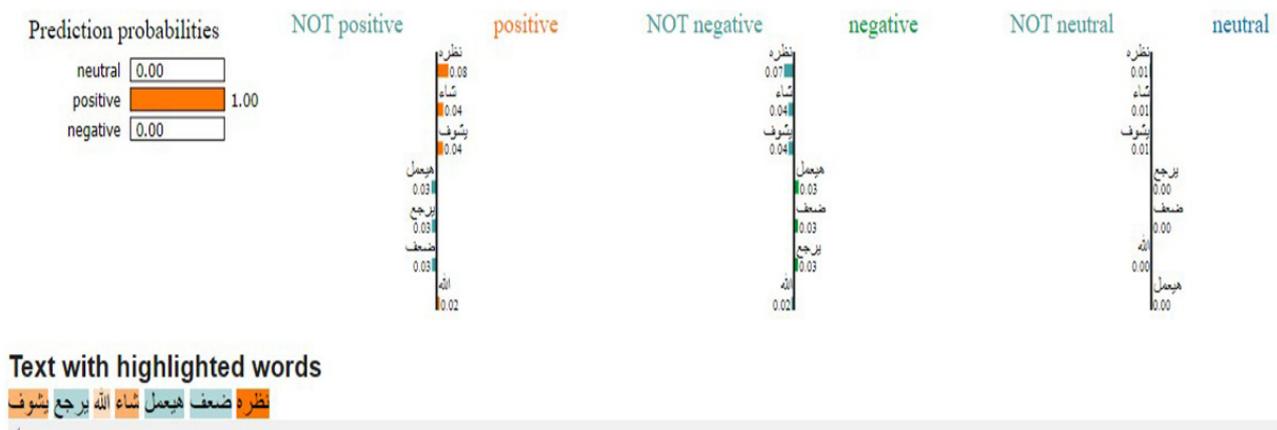


Figure 4. LIME results for LASIK surgery Positive sentiment.

Meanwhile, Figure 5 shows an illustration of a sentence that was originally classified as having neutral sentiment, as it is initially a sentence regarding what a clinic offers; the

text states “Offers on eye deficiency with the latest technologies, great price, free checkup and consultation”. The word “باحث” is classified as neutral, because it means “Latest” which can be used in a positive or negative context, based on its usage within this sentence. In this case, it is neutral, as it simply describes the latest technologies offered by the clinic. Other words, such as “الإبصار” and “لعلاج”, are used to represent eyesight and provide a cure for eyesight problems but are mentioned in a casual way to be read by users across social media. By applying XAI as a proposed approach, words and their importance can be better indicated. For example, the words used within the LASIK surgery data set can be assessed to further provide more useful insights about what concerns the potential patients may have before undergoing the surgery. These keywords may also be used as main search keywords within marketing campaigns used by clinics and hospitals, potentially leading to an increase in their reach to a wide range of potential patients. This can also be used to generate a safety index for the variations of LASIK surgeries, as determined by the experience of previous patients across Twitter users. Furthermore, we compared the count of words that appeared in the sentiments in Figure 5 with the probability weights of the words in Table 3, which indicated that some of the word counts were low. Words such as “باحث” (which means “latest”) had a larger probability of 0.46 and a word count of 12 times, while “الإبصار” (“eyesight”) had a probability of 0.18 and a word count of 21. Even though the second word had a higher word count, a higher probability was given to the word “latest” due to its usage across sentences specifically emphasizing the latest technologies used. Furthermore, words such as “فحص” and “التقنيات” had the same word count but different weights, due to the main subject of the sentence itself being eyesight.



Figure 5. LIME results for LASIK surgery neutral sentiment.

Table 3. Word counts and LIME-based weights of words.

Word	Weight	Word Count
الإبصار	0.18	21
باحث	0.46	12
لعلاج	0.20	9
فحص	0.15	18
التقنيات	0.09	18

While some words had a lower word occurrence, LIME also gave them significance according to their appearance with respect to the single sentiment itself, considering the occurrence of several variants of the same word with the same meaning. Therefore, LIME elaborates and works significantly well with Arabic text. According to [1], the utilized LIME XAI method for LASIK surgeries satisfied two out of the seven purposes stated as the main reasons for XAI applications: transparency, allowing users and decision-makers to further apply compatible decisions, and reliability, regarding the attention-based LSTM model, which can be proven according to the model performance on the data set.

Finally, an example of a Negative sentiment within the applied case study is shown in Figure 6. The sentiment involves the statement of the regret of a user after having the LASIK surgery: “I should go back to my eyeglasses”. The words “ارجع” (“go back”), “النضارة” (“eyeglasses”), and “لازم” (“Must”) are considered indicators of a negative experience here regarding the surgery, which made the user consider going back to wearing eyeglasses. Even though the words equivalent to “Must” and “go back” could have a positive impact within the Arabic language itself, according to the context of this particular sentiment, LIME was able to emphasize that these words were the main reason for the negative classification of this sentiment.

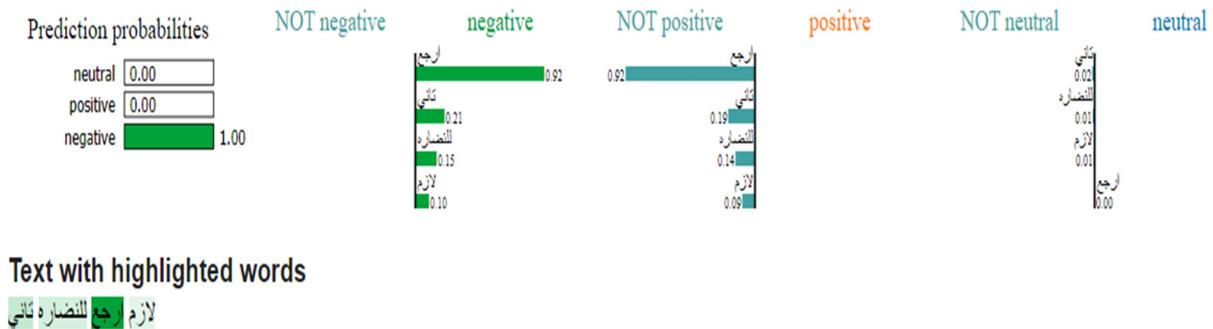


Figure 6. LIME results for LASIK surgery negative sentiment.

### 5. Discussion

An evaluation of the LSTM on the proposed data set was conducted, due to its size and nature. We measured the Recall, Precision, F1-score, and Accuracy. Table 4 shows the results across the data set, indicating an accuracy of 79.1%. As the data were unbalanced, the F1-score is a valuable metric to take into consideration, which reached 0.71 for the model. This can be considered a promising result, considering the nature of the data set. The recall measures the extent to which the model correctly classifies sentiment polarities. For our model and data set, the recall reached 0.76. On the other hand, the precision reached 0.71, which is also a promising result. Finally, Figure 7 illustrates how the model performed across the epochs within training and validation phases. While it achieved high accuracy in the Arabic text data set, we did not only aim for accuracy, but also good justification performance, regarding how well the model classified sentiments into negative, positive, and neutral.

Table 4. Evaluation Results.

Data Set	Accuracy	Precision	Recall	F1-Score
[23]	79.1%	0.71	0.76	0.71

In this work, we demonstrated several experiments following on from our previous work [8], which showed that the attention-based model had the best performance in word-level ASA. Furthermore, due to the complexity of the LSTM as well as the Arabic text, here we aimed to further justify how the attention-based LSTM classified the sentiments within Arabic text. We demonstrated our work on a domain-specific Arabic data set. First, we created a domain-specific data set for LASIK surgery feedback across Twitter users. The data set was labeled manually by two annotators. Subsequently, the proposed data set went through pre-processing and feature selection.

We then applied the XAI LIME method, due to its great performance across different studies in various fields, which led us to further examine its potential regarding Arabic, due to its high complexity and variations. In this study, we observed that in the application of ASA with the help of an attention-based LSTM and using LIME as a post hoc explanation method, we could determine the sentiment classification based on specific words within

the context of LASIK surgery Arabic texts; as such, we could conclude that LIME works well in the face of the complexity of Arabic text, especially with respect to its various dialects used across social media services (which led to challenges and considerations in the labelling and pre-processing steps). For this study, many trials were carried out, with the consideration of keeping the collected text closer to what was originally written by users. Therefore, not applying normalization and not removing repetitive characters were essential points that helped to improve the performance of both the attention-based model and LIME, as well as how words were classified according to their importance, regardless of the variation in the same words. In this line, we presented specific examples for the sake of explaining how the DL model classified the sentiments across the Arabic text. Second, we aimed to gain further insights into the main concerns of potential LASIK surgery patients, which could be helpful in developing a safety index for a future marketing campaign or another targeted promotion approach for future potential patients. Finally, the LIME results were promising, in terms of both presented examples of positive and neutral tweets. We then demonstrated a comparison and described how LIME classified words according to their significance within the sentiment analysis context, indicating that it works well both within the domain-specific Arabic text data set as well as for further evaluation of the Attention-based LSTM model across the domain-specific LASIK surgery Arabic text data set [26].

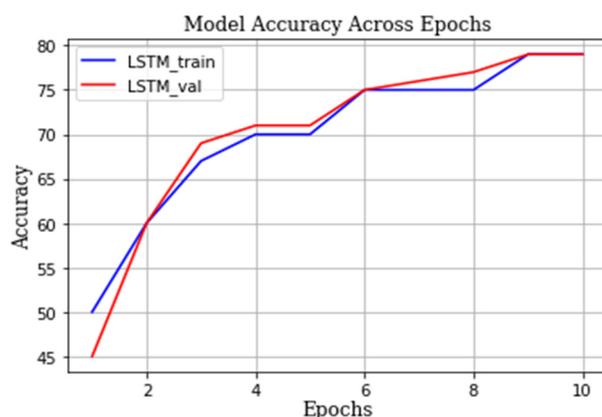


Figure 7. Performance Assessment.

## 6. Conclusions

In this study, an LSTM approach based on an attention layer and word count in the embedding layer was applied. For further analysis and comprehension of results, LIME (which is an XAI method) was applied to the attention-based LSTM model, even though it did not reach the expected accuracy. We were able to achieve an accuracy of 79.1%, which can be considered good due to the complexity and nature of Arabic text. The end goals here were primarily to further explain the classification of sentiments by the DL model, as they are considered black-box models. We have confirmed that the previously mentioned attention-based LSTM Model performed the best across different data sets in a previous work [8]. Subsequently, we used this model on a domain-specific data set composed of the opinions of LASIK surgery patients across Twitter, in order to clarify how the sentiments were classified into the corresponding classes, which were proposed based on word count. Furthermore, we applied LIME across three examples relating to the three sentiments in Arabic text about LASIK surgery, in order to further understand the concerns of the patients when trying to commit to an eye surgery based on Twitter posts, as well as how these words were output with their corresponding probabilities. This, in turn, is expected to help in choosing better keywords when targeting patients in future marketing campaigns, which may lead to a higher rate of coverage of an event. We deduced that LIME works well concerning Arabic text, due to its nature of checking the words within a local sentiment, according to the complex morphology of Arabic language and the variant dialects used

across users, where a word can have different sentimental impact depending on where it is placed within the Arabic sentence. These results can help in further trusting the outcomes presented by deep learning models for non-expert users and decision-makers. Finally, our future work will consist of applying XAI methods to multi-dialect Arabic data sets, which are considered challenging due to the variety and variations of words with exact meanings, in order to evaluate how LIME works in a multi-dialect data set as well as considering several other sentiment levels, such as character- and document-level.

**Author Contributions:** Conceptualization, M.K. and A.A.H.S.; Supervision, Y.A.; methodology, software, validation, formal analysis, and visualization, Y.A.; data curation, Investigation, writing—original draft preparation. M.K., A.A.H.S. and Y.A.; writing—review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data set publicly available via Abdelfattah Youmna. [lasikSurgery-arabic-text-dataset.Kaggle.com.10.34740/kaggle/dsv/42722722022](https://www.kaggle.com/datasets/younnahabdelmattah/lasik-surgery-arabic-text-dataset) (accessed on 1 October 2022). Available from <https://www.kaggle.com/datasets/younnahabdelmattah/lasik-surgery-arabic-text-dataset> (accessed on 1 October 2022).

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

## Abbreviations

Abbreviation	Terminology
DL	Deep learning
ML	Machine Learning
LIME	Local Interpretable Model-agnostic Explanations
LSTM	Long Short-Term Memory
SHAP	SHapley Additive exPlanations
ASA	Arabic Sentiment Analysis

## References

1. Fiok, K.; Farahani, F.V.; Karwowski, W.; Ahram, T. Explainable artificial intelligence for education and training. *J. Def. Model. Simul. Appl. Methodol. Technol.* **2021**, *19*, 133–144. [CrossRef]
2. Arrieta, A.B.; Díaz-Rodríguez, N.; del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
3. Feldman, R. Techniques and applications for sentiment analysis. *Commun. ACM* **2013**, *56*, 82–89. [CrossRef]
4. 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]
5. Kim, H.; So, K.K.; Wirtz, J. Service robots: Applying social exchange theory to better understand human–robot interactions. *Tour. Manag.* **2022**, *92*, 104537. [CrossRef]
6. Oueslati, O.; Cambria, E.; HajHmida, M.B.; Ounelli, H. A review of sentiment analysis research in Arabic language. *Future Gener. Comput. Syst.* **2020**, *112*, 408–430. [CrossRef]
7. Aljameel, S.S.; Alabbad, D.A.; Alzahrani, N.A.; Alqarni, S.M.; Alamoudi, F.A.; Babili, L.M.; Aljaafary, S.K.; Alshamrani, F.M. A sentiment analysis approach to predict an individual’s awareness of the precautionary procedures to prevent COVID-19 outbreaks in Saudi Arabia. *Int. J. Environ. Res. Public Health* **2021**, *18*, 218. [CrossRef] [PubMed]
8. Abdelwahab, Y.; Kholief, M.; Sedky, A. An experimental survey of ASA on DL classifiers using multi-dialect arabic texts. In Proceedings of the Future of Information and Communication Conference 2022, San Francisco, CA, USA, 3–4 March 2022.
9. Alaff, A.J.; Mukhairez, H.H.; Kose, U. An explainable artificial intelligence model for detecting COVID-19 with twitter text classification: Turkey case. In Proceedings of the International Conference on Computing and Communication Systems 2021, Shillong, India, 28–30 April 2020; Springer: Singapore, 2020; pp. 87–97.
10. Rathore, R.K.; Kolonin, A. Explorative study of explainable artificial intelligence techniques for sentiment analysis applied for english language. In Proceedings of the International Conference on Paradigms of Communication, Computing and Data Sciences 2022, Kurukshetra, India, 7–9 May 2021; Springer: Singapore, 2021; pp. 861–868.
11. Gite, S.; Khatavkar, H.; Srivastava, S.; Maheshwari, P.; Pandey, N. Stock prices prediction from financial news articles using LSTM and XAI. In Proceedings of the Second International Conference on Computing, Communications, and Cyber-Security 2021, Delhi, India, 3–4 October 2020; Springer: Singapore, 2020; pp. 153–161.

12. Adak, A.; Pradhan, B.; Shukla, N.; Alamri, A. Unboxing deep learning model of food delivery service reviews using explainable artificial intelligence (XAI) technique. *Foods* **2022**, *11*, 2019. [[CrossRef](#)] [[PubMed](#)]
13. Chowdhury, K.R.; Sil, A.; Shukla, S.R. Explaining a black-box sentiment analysis model with local interpretable model diagnostics explanation (LIME). In *Advances in Computing and Data Sciences, Proceedings of the 5th International Conference on Advances in Computing and Data Sciences, Nashik, India, 23–24 April 2021*; Springer: Cham, Switzerland, 2021; pp. 90–101.
14. Kumar, A.; Dikshit, S.; Albuquerque, V.H. Explainable artificial intelligence for sarcasm detection in dialogues. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 2939334. [[CrossRef](#)]
15. Choi, I.H.; Kim, Y.S.; Lee, C.K. A Study of the classification of IT jobs using LSTM and LIME. In *Proceedings of the 9th International Conference on Smart Media and Applications, Jeju, Korea, 17–19 September 2020*; pp. 248–252.
16. Tang, G.; Zhang, L.; Yang, F.; Meng, L.; Cao, W.; Qiu, M.; Ren, S.; Yang, L.; Wang, H. Interpretation of learning-based automatic source code vulnerability detection model using LIME. In *Knowledge Science, Engineering and Management, Proceedings of the International Conference on Knowledge Science, Engineering and Management, Tokyo, Japan, 14–16 August 2021*; Springer: Cham, Switzerland, 2021; pp. 275–286.
17. Aporna, A.A.; Azad, I.; Amlan, N.S.; Mehedi, M.H.; Mahub, M.J.; Rasel, A.A. Classifying offensive speech of bangla text and analysis using explainable AI. In *Advances in Computing and Data Sciences, Proceedings of the 6th International Conference on Advances in Computing and Data Sciences, Kurnool, India, 22–23 April 2022*; Springer: Cham, Switzerland, 2022; pp. 133–144.
18. Pérez-Landa, G.I.; Loyola-González, O.; Medina-Pérez, M.A. An Explainable Artificial Intelligence Model for Detecting Xenophobic Tweets. *Appl. Sci.* **2021**, *11*, 10801. [[CrossRef](#)]
19. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why should i trust you? ” Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016*; pp. 1135–1144.
20. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.
21. Malandri, L.; Mercorio, F.; Mezzanzanica, M.; Nobani, N.; Seveso, A. ContrXT: Generating contrastive explanations from any text classifier. *Inf. Fusion* **2022**, *81*, 103–115. [[CrossRef](#)]
22. Rajagopal, D.; Balachandran, V.; Hovy, E.; Tsvetkov, Y. Selfexplain: A self-explaining architecture for neural text classifiers. *arXiv* **2021**, arXiv:2103.12279.
23. Guidotti, R. Evaluating local explanation methods on ground truth. *Artif. Intell.* **2021**, *291*, 103428. [[CrossRef](#)]
24. Alsabaani, N.; Alshehri, M.S.; AlFlan, M.A.; Awadalla, N.J. Prevalence of laser refractive surgery among ophthalmologists in Saudi Arabia. *Saudi J. Ophthalmol.* **2020**, *34*, 116. [[CrossRef](#)] [[PubMed](#)]
25. Boyd, K. LASIK—Laser Eye Surgery. American Academy of Ophthalmology. 23 August 2022. Available online: <https://www.aaao.org/eye-health/treatments/lasik> (accessed on 23 August 2022).
26. Abdelfattah Youmna. LasikSurgery-Arabic-Text-Dataset. *Kaggle.com*. 10.34740/kaggle/dsv/4272272. (Dataset). 2022. Available online: <https://www.kaggle.com/datasets/youmnaabdelfattah/lasik-surgery-arabic-text-dataset> (accessed on 1 October 2022).
27. TensorFlow Core v2.9.1. TensorFlow. Available online: [https://www.tensorflow.org/api\\_docs/python/tf/keras/preprocessing/text/Tokenizer](https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/text/Tokenizer) (accessed on 7 June 2022).
28. Local Interpretable Model-Agnostic Explanations (Lime)—Lime0.1 Documentation. Available online: <https://lime-ml.readthedocs.io/en/latest/> (accessed on 27 October 2022).