

## SENTIMENT ANALYSIS BASED ON MACHINE LEARNING METHODS ON TWITTER DATA USING oneAPI

Fatih Şengül<sup>1,\*</sup>, Kemal Adem<sup>2</sup> and Esra Kavalcı Yılmaz<sup>3</sup>

<sup>1,3</sup>Dept. of Defense Technologies, Sivas University of Science and Technology, Turkey

<sup>2</sup>Dept. of Computer Engineering, Sivas University of Science and Technology, Turkey

\*(210102002@sivas.edu.tr) Email of the corresponding author

**Abstract** – With the gradual development of technology, the use of digital social networks has become widespread. As a result of this increased use, the information sharing process has accelerated and increased. This has created an enormous amount of data on digital social networks. With the increasing amount of data, it has made it possible to make inferences about people, products, companies and many other areas. Various fields of study have emerged to process and analyze the data. One of these fields of study is sentiment analysis. Sentiment analysis is the analysis of data from different sources such as text, sound or image by classifying the attitude towards a subject as positive, negative or neutral. In this study, sentiment analysis was performed using Sentiment140 dataset created through Twitter application, one of the digital social network platforms. After preprocessing on the dataset containing English Tweet messages, Bernoulli Naive Bayes, Linear Support Vector Machine, Logistic Regression and artificial neural network algorithms LSTM and CNN methods were used in a hybrid way. Classification results were evaluated with f1 score. Accuracy rates of 0.80%, 0.82%, 83%, 85% were achieved for Bernoulli Naive Bayes, Linear Support Vector Machine, Logistic Regression, LSTM and CNN classifications respectively.

**Keywords** –Sentiment Analysis, Twitter Data, oneAPI

### I. INTRODUCTION

With the development of internet technologies, the popularity of digital social networks has been increasing. Digital social networks have rapidly become popular as a new alternative to communication tools such as newspapers, television, and radio. Digital social network applications such as Facebook and Twitter, which are frequently used today, have accelerated the processes such as the formation, dissemination, and discussion of information. With the development of digital social networks, one-sided news and information sharing has become interactive. Users of digital social networking platforms can both share information and show reactions such as instant comments and likes. Users can instantly share the development in any subject or news on their digital network. As in many areas of technology, social

networking technologies have brought various risks and opportunities to companies and individual users at the same time. Organizations, communities, associations, and governments shape their policies according to the feedback from digital social networks. Users can easily express their feelings and thoughts on a topic through posts on digital social networks. Twitter, one of the most popular digital social networking platforms today, is used worldwide. Twitter shares 347,200 tweets per minute [1]. This data reflects the enormous amount of public data generated by the increasing use of social networking platforms. The Twitter social networking platform allows users to post tweets of up to 140 characters in length. Each tweet has Reply, Retweet and Like actions. These features can be used to find out how popular a tweet is and how many people it has reached. From the information

about who users are communicating with, various analyses can be made, such as the source of the original information, the way the information spreads, and identifying influential users. One of the analyses used in this sense is sentiment analysis. Sentiment analysis is the determination of whether an article is positive or negative. Sentiment analysis can be used in many areas such as market research for companies, early detection of terrorist incidents or provocations against the state for governments and intelligence services [2].

Sentiment analysis refers to the computational process of detecting and classifying opinions expressed within a text, aiming to identify and categorize them. [3]. Sentiment analysis, also known as opinion mining, is an active area of research that focuses on the identification of subjective information, including emotions, opinions, and attitudes expressed within a text. This field utilizes methods and techniques from diverse disciplines like natural language processing, statistics, and computer science [4]. Modeling sentiment analysis can be approached as a text classification problem [5]. Based on the level of granularity in the classification process, sentiment analysis can be classified into three main categories: document-level sentiment analysis, sentence-level sentiment analysis, and entity/feature-level sentiment analysis [6]. Document-level sentiment analysis involves determining the overall sentiment expressed in a text document, while phrase-level sentiment analysis focuses on identifying the sentiment associated with specific phrases within the document. Entity/feature-level sentiment analysis aims to ascertain the sentiment towards particular features of an entity. In sentiment analysis, the methods used can be broadly categorized into two main classes: machine learning-based methods and dictionary-based methods [6]. In machine learning based sentiment analysis methods, machine learning classifiers are trained with the opinion polarity labeled data set to create a classification model, which is then used to determine the opinion polarity of new examples. In dictionary-based sentiment analysis methods, sentiment analysis is performed by creating a dictionary containing opinion words [7]. In text mining, decision tree algorithms, statistics-based classifiers (such as Naive Bayes algorithm), linear classifiers (such as support vector machines) and artificial neural networks have been successfully

applied [8]. In order to apply machine learning classifiers in text classification, it is first necessary to extract and select text features using an appropriate method. Representation methods such as term frequency, term presence, phrase elements, opinion words are frequently used representation structures [6].

There are many studies in the literature on sentiment analysis using machine learning classifiers on Twitter data. In their study in 2022, Tan et al. achieved 89.81% accuracy rate on Sentiment140 dataset with the hybrid models they created using RoBERTa, LSTM and BiLSTM methods [9]. Kamyab et al. used Sentiment 140 dataset in their study conducted in 2021. As a result of their experiments using CNN and Bi-LSTM models, they achieved 87% accuracy rate on Sentiment 140 dataset [10]. Al-deen et al. [11] achieved the highest accuracy rate of 82% on Sentiment 140 dataset using DNN-MHAT model [11]. Gaye et al. reached 99% accuracy rate on Sentiment 140 dataset using LSTM method in their study in 2021[12]. In the study by Goel et al. Sentiment140 dataset was used in the training phase. In the study where they used Naive Bayes method, they tested with 100 Tweet data and achieved an accuracy rate of 58.40% [13]. Ayan et al. used Twitter API to classify Tweet messages with Islamophobic content. They used Linear Ridge and Naive Bayes methods. In Ridge Regression, 96.3% and 95.3% correct results were obtained in Naive Bayes Classifier [14]. In their study conducted in 2021, Habib et al. compared machine learning methods for classification using Sentiment140 dataset. BOW, TF-IDF, doc2vec and word2vec methods were used for feature extraction. At the end of the study, they proposed the most successful method as Logistic regression with 82.76% accuracy rate [15].

With the healthy analysis of large data pools on digital social networking platforms, people's emotional states and analysis, positive and negative opinions on a subject can be determined. Within the scope of this study, the classification performance of machine learning and deep learning methods were compared using Sentimet140 dataset containing English Twitter messages. The text contents in the dataset were first preprocessed and then the classification process was performed. The model with the highest classification performance was proposed as a decision support system.

In the rest of the paper, Section-2 Materials and methods provides information about the Sentiment140 dataset used in this study and explains the classification methods used. Section-3 Results presents a comparison of the values obtained with the classification methods used in the study. Section-4 discussion shows the results of all analyses and discussions. Conclusion and future work are given in Section-5.

## II. MATERIALS AND METHOD

In this section of the study, the dataset and the classification methods used are explained in detail.

### A. Dataset

In order to perform sentiment analysis on English Twitter messages, Sentiment140 dataset created by Stanford University was used. The dataset, created using the Twitter API, contains 1,600,000 tweets. It consists of 6 fields in total: polarity of the tweet messages (0 = negative, 4 = positive), id (Tweet ID), date the tweet message was shared, username of the tweeter on the Twitter application, content of the tweet message and flag value. An example of the content of the dataset is shown in Figure 1.

In the dataset, 1.6 million Tweet messages are categorized into two classes: positive and negative content. The distribution in the dataset is shown in Figure 2.

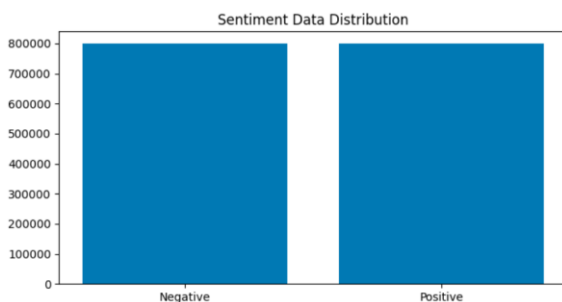


Fig. 2 Data set distribution

### B. Methods

Since the data obtained using the Twitter API via the Twitter application is unstructured, some preprocessing on Tweet messages is required for sentiment analysis. The inclusion of additional information in tweets, such as hashtags, URLs, emoji, and symbols, necessitates the utilization of effective data cleaning techniques to enhance the performance of machine learning algorithms. As natural language processing (NLP) models are unable to process this type of data directly, it becomes imperative to preprocess tweets by removing irrelevant tokens and retaining only those that carry meaningful information for the model. Additionally, as part of the preprocessing step, all text characters are converted to lowercase. Then the link information starting with "http", "https" or "www" is replaced with "URL". At the same time, the username information that will not be used as an attribute for sentiment analysis is changed to '<user>'. In order to perform targeted analysis by parsing meaningful words, 3 or more consecutive letters are replaced with 2 letters (e.g., 'Heyyy' to 'Heyy'). Emoji emoticons used in tweets are also tagged as '<smile>'. English expressions were used in the dataset, and expressions such as "can't" in English were changed to "cannot" in order to better apply NLP algorithms. Non-character data such as numbers and symbols in tweet messages were replaced with spaces. The preprocessing steps are crucial for the analysis and the order in which they are performed is also important. For example, if punctuation marks are removed before replacing URLs, it may result in URLs not being detected. In order to classify Positive and Negative tweets from our dataset, we determine which words occur the most. After doing this, a word cloud is created using the World Cloud library. The resulting positive and negative word clouds are as shown in Figure 3 and Figure 4 respectively.

	sentiment	id	date	query	user_id	text
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by ...
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Man...
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
4	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	Karoli	@nationwideclass no, it's not behaving at all...

Fig. 1 Content example of the dataset

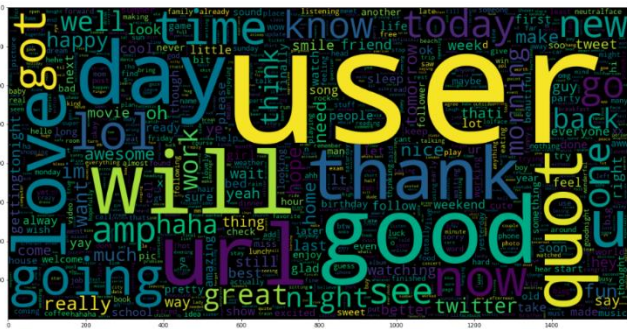


Fig. 3 Positive word cloud in the dataset



Fig. 4 Positive word cloud in the dataset

C. Classifiers

In the study, five different classification algorithms were used to determine the sentiment poles of Twitter data. These are Bernoulli Naive Bayes, Linear support vector machine, Logistic Regression and LSTM and CNN method from artificial neural network algorithms. LSTM and CNN model were used as hybrid.

- Naive Bayes

It is a classification technique based on Bayes' theorem. Naive Bayes classification assumes that there are certain features in a class that are unconnected to the structure of other features. Even if these features are connected to others or depend on each other's presence, they all contribute to independent probabilities. The naive bayes model is simple in terms of setup and has been used especially for very large datasets [16].

- Linear Support Vector Machine

The primary objective of a linear support vector machine algorithm is to identify a hyperplane that can accurately classify data points within an n-dimensional space (where n represents the number of features). When attempting to separate two classes of data points, numerous hyperplanes can be

considered. The underlying principle involves identifying a plane that maximizes the distance between the data points belonging to each class [17].

- Logistic Regression

The logistic regression method is employed to make predictions on a categorical dependent variable based on a provided set of independent variables. The logistic regression method uses a sigmoid function that estimates two maximum values [16].

- Convolutional Neural Networks (CNN)

In recent years, Convolutional neural networks (CNNs) have been frequently used in image processing, text processing and audio processing. These properties of CNNs are independent of the number of dimensions [18]. While CNNs are widely used in computer vision, they have also been applied to natural language. A convolutional neural network for text operates in only two dimensions and filters only need to act in the time dimension [19], [20]. Figure 5 shows the structure of the CNN architecture.

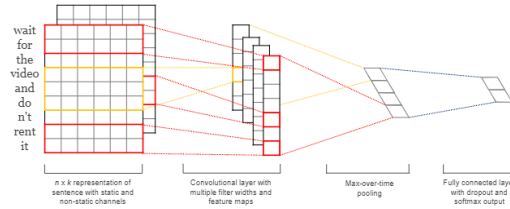


Fig. 5 The structure of the CNN architecture

- Long Short Term Memory

LSTM, which stands for Long Short Term Memory, is a recurrent neural network that has gained significant popularity in the machine learning domain. Initially introduced by Hochreiter and Schmidhuber in 1997, LSTM and its subsequent variations have undergone further advancements through the collaborative efforts of numerous researchers. As a result, these LSTM variants have gained widespread adoption and usage [21]. LSTM sets itself apart from conventional feed-forward neural networks by incorporating feedback connections. This unique characteristic allows LSTM to handle not only instantaneous data like images but also sequential data such as speech or video. Practical applications of LSTM encompass various tasks, including unsegmented, connected handwriting recognition, speech recognition, as well



as detecting anomalies in network traffic or intrusion detection systems (IDSs). Typically, an LSTM unit comprises a cell, an entry gate, an exit gate, and a forget gate. The cell remembers values at variable length time intervals and these three gates regulate the flow of information in and out of the cell [22], [23]. LSTM networks are a suitable method for classification and prediction based on time series data.

### III. RESULTS

In order to make better use of machine learning methods in the classification process and to obtain better results, the messages in the data set were subjected to preprocessing. After the data was preprocessed, the classification phase was started. Four different algorithms were used in the classification phase. Accuracy, Precision, Recall (Sensitivity) and F score values were used to evaluate the success of the algorithms. The F1 score is the harmonic mean of the precision and sensitivity values. Since it is a measurement metric that includes not only false negatives or false positives but also all error costs, the f1 score was used [16]. In equations (4), (5), (6), (7) used in the performance comparison of the models, TP, TN, FP, FN denote true positive, true negative, false positive, false negative values, respectively.

Table 1. Complexity Matrix

	Positive Prediction	Negative Prediction
True Positive	True Positive (TP)	False Negative (FN)
True Negative	False Positive (FP)	True Negative (TN)

#### Accuracy

Accuracy is the number of all correct predictions divided by the entire dataset (4).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

#### Precision

Precision is the number of correct positive predictions divided by the number of all positive predictions (5).

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

#### Recall

The number of true positive predictions divided by the number of true positives in the dataset (6) is called recall.

$$Recall = \frac{TP}{FN} \quad (6)$$

#### F1 Score

The F score is used to show the trade-off between sensitivity and recall. The F score is obtained by equation (7).

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

In this study, Intel DevCloud cloud system was used for preprocessing Twitter messages and for training and testing machine learning models [24]. Intel oneAPI framework is used for faster machine learning and deep neural network model training. oneAPI platform allows the developed codes to utilize multiple hardware architectures such as multi-core CPUs, GPUs or other hardware using a single source [25]. oneAPI framework increases the processing speed by effectively utilizing the parallel computing power of multi-core processors and graphics processing units (GPUs). It also provides the ability to utilize hardware resources with a single source code, making it easier for programs to run on different hardware platforms and increasing portability. This allows developers to optimize their applications on a wider range of hardware [26]. Thus, it shortened the experiment times considerably. After preprocessing the Twitter data, the Bernoulli NB method, one of the machine learning methods, was applied. Accuracy, Precision, Sensitivity and F-1 Measure values of the applied model are as shown in Table 2.

Table 2. Results obtained in classification with Bernoulli Naive Bayes method

Class	Precision	Recall	F1 Score
Negative	0.81	0.79	0.80
Positive	0.80	0.81	0.80

When the results of the Bernoulli Naive Bayes method are evaluated, it is seen that all classification processes are close to each other. The overall accuracy rate was 80% as shown in Table 2. The binary classification model evaluations created with Linear Support Vector Machine, another machine learning method, are as shown in Table 3.

Table 3. Results obtained in classification with linear support vector machine method

<b>Class</b>	<b>Precision</b>	<b>Recall</b>	<b>F1 Score</b>
Negative	0.82	0.81	0.82
Positive	0.81	0.83	0.82

As a result of the classification process performed with the linear support vector machine, negative and positive classifications are equal, and the classification rate of the model is 82%. Another model created to perform binary classification on the preprocessed data set is the Logistic Regression method.

Table 4. Results obtained in classification with logistic regression method

<b>Class</b>	<b>Precision</b>	<b>Recall</b>	<b>F1 Score</b>
Negative	0.83	0.82	0.83
Positive	0.82	0.84	0.83

As can be seen in Table 4, the Logistic Regression method achieved an accuracy rate close to the classification methods given before (%83). In order to perform sentiment classification on Twitter messages, many different methods were tried with different parameters. After machine learning methods, a hybrid model was created using artificial neural networks. The hybrid model has an embedding layer that converts positive integers into a fixed-size vector, two LSTM layers, a CNN and MaxPooling layer, a dense and a Softmax classification layer.

Table 5. Classification results obtained with LSTM and CNN

<b>Class</b>	<b>Precision</b>	<b>Recall</b>	<b>F1 Score</b>
Negative	0.84	0.85	0.85
Positive	0.85	0.84	0.84

When the results obtained with the hybrid model are evaluated, it has reached 85% classification accuracy. As can be seen in Table 5, it is seen that the training process of the hybrid artificial neural network model takes a lot of time compared to other methods, although it gives a successful result in the classification process compared to other models.

#### IV. DISCUSSION

The hybrid neural network model is more successful than other methods due to factors such as more comprehensive feature representation, depth and flexibility, capturing long-term dependencies, and the ability of CNN to capture local features. By providing a more comprehensive feature representation by considering the word order in the text, it more accurately captures the emotional content of words and phrases in the text. In addition, since it has a deeper structure thanks to the layers of LSTM and CNN, it can better capture long-term dependencies and better detect local features thanks to CNN.

#### V. CONCLUSION

Sentiment analysis has been used extensively by companies, organizations, and governments in recent years. The results obtained provide important inferences for companies in their marketing processes and for governments to evaluate the reaction of the society on any issue. Sentiment analysis can be from different sources such as text and audio. In this study, sentiment classification of Twitter messages received through Twitter application was performed using Sentiment140 dataset. Machine learning and deep learning-based models are compared for the classification process. The tweet messages in the dataset were classified as either positive or negative. Although the models compared were very close to each other, the hybrid artificial neural network model trained and tested with Sentiment140 dataset was found to be the most successful model with 85% accuracy rate.

Future studies could use larger and more diverse datasets and focus on different languages and cultures. It could also focus on optimizing deep learning models and developing deep learning models with more advanced emotional inference capabilities.

#### ACKNOWLEDGMENT

We would like to express our sincere gratitude to Intel for providing us with the Intel oneAPI toolkit, which played a crucial role in our research work. The powerful features and capabilities of Intel oneAPI greatly contributed to the success of our study in sentiment analysis. We also extend our appreciation to Intel for granting us access to the

DevCloud cloud system, which enabled us to conduct efficient training and testing stages for our models. The availability of DevCloud was instrumental in achieving accurate and reliable results. We are grateful for the support and resources provided by Intel, which significantly enhanced the quality and effectiveness of our work in utilizing the Intel oneAPI toolkit and leveraging the DevCloud cloud system for our research in sentiment analysis.

## REFERENCES

- [1] E. Cano-Marin, M. Mora-Cantalops, S. Sánchez-Alonso, "Twitter as a predictive system: A systematic literature review," *Journal of Business Research*, Volume 157, 2023.
- [2] M. Albayrak, K. Topal and V. Altıntaş, "Sosyal medya üzerinde veri analizi: Twitter," *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Kayfor 15 Özel Sayısı, 1991-1998.
- [3] R. N. Behera, S. Dash and M. Roy, "A Novel Machine Learning Approach for Classification of Emotion and Polarity in Sentiment140 Dataset", *Conference: 3rd International Conference on Business & Information Management*, 2016.
- [4] A. Onan, "Twitter Mesajları Üzerinde Makine Öğrenmesi Yöntemlerine Dayalı Duygu Analizi", *Yönetim Bilişim Sistemleri Dergisi*, c. 3, sayı. 2, ss. 1-14, Ara. 2017.
- [5] A. Onan and S. Korukoğlu, "A feature selection model based on genetic rank aggregation for text sentiment classification", *Journal of Information Science*, 2023.
- [6] W. Medhat, A. Hassan, and H. Korash, "Sentiment analysis algorithms and applications: A survey." *Ain Shams engineering journal* 5.4, 2014.
- [7] A. Onan, and S. Korukoğlu, "Makine öğrenmesi yöntemlerinin görüş madenciliğinde kullanılması üzerine bir literatür araştırması", *Pamukkale University Journal of Engineering Sciences* 22.2, 2016.
- [8] C.C. Aggarwal, and C. X. Zhai, "A survey of text classification algorithms." *Mining text data*, 2012.
- [9] K.L. Tan, C.P. Lee, K.M. Lim, K.S.M. Anbananthen, "Sentiment analysis with ensemble hybrid deep learning model." *IEEE Access* 10, 2022.
- [10] M. Kamyab, G. Liu, and M. Adjeisah. "Attention-based CNN and Bi-LSTM model based on TF-IDF and glove word embedding for sentiment analysis." *Applied Sciences* 11.23, 2021.
- [11] S. Al-deen, H. Saleh, Z. Zeng, R. Al-sabri, A. Hekmat "An improved model for analyzing textual sentiment based on a deep neural network using multi-head attention mechanism." *Applied System Innovation* 4.4, 2021.
- [12] B. Gaye, D. Zhang, and A. Wulamu, "A tweet sentiment classification approach using a hybrid stacked ensemble technique." *Information* 12.9, 2021.
- [13] A. Goel, J. Gautam, and S. Kumar, "Real time sentiment analysis of tweets using Naive Bayes." *2nd International Conference on Next Generation Computing Technologies (NGCT)*. IEEE, 2016.
- [14] B. Ayan, B. Kuyumcu and B. Ceylan, "Twitter Üzerindeki İslamofobik Twitlerin Duygu Analizi ile Tespiti." *Gazi University Journal of Science Part C: Design and Technology* 7.2, 2019.
- [15] M.W. Habib and Z.N. Sultani, "Twitter sentiment analysis using different machine learning and feature extraction techniques." *Al-Nahrain Journal of Science* 24.3, 2021.
- [16] E. Kumaş, "Türkçe Twitter Verilerinden Duygu Analizi Yapılırken Sınıflandırıcıların Karşılaştırılması." *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi* 2.2, 2021.
- [17] R. Gandhi, "Support Vector Machine — Introduction to Machine Learning Algorithms," *Towards Data Science*, 2018.
- [18] K. Adem and S. Kiliçarslan. "COVID-19 diagnosis prediction in emergency care patients using convolutional neural network." *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi* 21.2, 2021.
- [19] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences." *arXiv preprint arXiv:1404.2188*, 2014.
- [20] Y. Chen, "Convolutional neural network for sentence classification." *MS thesis*. University of Waterloo, 2015.
- [21] M. Turkoglu, D. Hanbay, and A. Sengur. "Multi-model LSTM-based convolutional neural networks for detection of apple diseases and pests." *Journal of Ambient Intelligence and Humanized Computing*, 2019.
- [22] H. Sak, A.W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling.", 2014.
- [23] X. Li, and X. Wu, "Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition." *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015.
- [24] Intel, "Intel DevCloud," 2019.
- [25] S. Christgau, and T. Steinke, "Porting a legacy cuda stencil code to oneapi." *2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. IEEE, 2020.
- [26] P.A. Martínez, B. Peccerillo, S. Bartolini, J. M. García, and G. Bernabé, "Applying Intel's oneAPI to a machine learning case study." *Concurrency and Computation: Practice and Experience* 34.13, 2022.