2022

M.Sc. in Electronics and Computer Engineering

Hatice Meltem NERGİZ

HASAN KALYONCU UNIVERSITY
GRADUATE SCHOOL OF
NATURAL AND APPLIED SCIENCES

MULTIMODAL CLASSIFIER FOR DISASTER RESPONSE

M. Sc. THESIS
IN
ELECTRONICS AND COMPUTER ENGINEERING

BY
Hatice Meltem NERGİZ ŞİRİN
Jan 2022

HASAN KALYONCU UNIVERSITY
GRADUATE SCHOOL OF
NATURAL AND APPLIED SCIENCES

MULTIMODAL CLASSIFIER FOR DISASTER RESPONSE

M.Sc. THESIS
IN
ELECTRONICS AND COMPUTER ENGINEERING

BY
Hatice Meltem NERGİZ ŞİRİN
2022

“Multimodal Classifier for Disaster Response”

M.Sc. Thesis

in

Electronics and Computer Engineering

Hasan Kalyoncu University

Supervisor

Asst. Prof. Dr. Saed ALQARALEH

By

Hatice Meltem NERGİZ ŞİRİN

2022

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES INSTITUTE M.Sc. ACCEPTANCE AND APPROVAL FORM

Electronics and Computer Engineering Department, Electrical and Electronics Engineering M.Sc. (Master of Science) program student **Hatice Meltem NERGİZ ŞİRİN** prepared and submitted the thesis titled **Multimodal Classifier for Disaster Response** defended successfully on the date of **14/01/2022** and accepted by the jury as an M.Sc. thesis.

Position	**Title, Name, and Surname Department/University**	**Signature**
Supervisor	Asst. Prof. Dr. Saed ALQARALEH Computer Engineering Department Hasan Kalyoncu University	
Jury Member	Asst. Prof. Dr. Mohammed Madi Computer Engineering Department Hasan Kalyoncu University	
Jury Member	Asst. Prof. Dr. Bülent HAZNEDAR Computer Engineering Deparment Gaziantep University	

This thesis is accepted by the jury members selected by the institute management board and approved by the institute management board.

Prof. Dr. ……

Director

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Hatice Meltem NERGİZ ŞİRİN

ABSTRACT

MULTIMODAL CLASSIFIER FOR DISASTER RESPONSE

NERGİZ ŞİRİN, HATİCE MELTEM

M.Sc. in Electronics and Computer Engineering.

Supervisor: Asst Prof. Dr. Saed ALQARALEH

Jan 2022, 67 pages

Nowadays, social media data can be used to make a huge difference in making correct decisions in time-critical situations, such as the event of natural disasters. Social media content consists of messages, images, and videos. In some cases, understanding the damage caused by natural disasters only from text is not enough in terms of analysis, the effect of disaster is better understood using visual data. Text datasets from social media platforms are widely used by researchers, and a limited number of studies have focused on the use of other content such as images. This is due to the fact that the number of tagged image datasets related to disasters is very limited.

Therefore, in this thesis, we aim to address this limitation by presenting a multimodal Turkish text and images dataset. Multimodal classification studies were carried out with the late fusion technique. Also, to achieve multimodal classification; a pre-trained LSTM model is used for classifying the text while a pre-trained CNN model is used for the visual content. Overall, concatenating both inputs in a multimodal learning architecture achieved an accuracy of 91.87%.

Key Words: Multimodal Classifier, Disaster Management, Tweet Text Classification, Tweet Image Classification, Turkish language

ÖZET

AFET MÜDAHALE İÇİN ÇOK MODLU SINIFLANDIRICI

NERGİZ ŞİRİN, HATİCE MELTEM

Yüksek Lisans Tezi, Elektronik Bilgisayar Müh. Bölümü

Tez Yöneticisi: Dr. Saed ALQARALEH

Ocak 2022, 67 Sayfa

Günümüzde sosyal medya verileri, doğal afet gibi zaman açısından kritik durumlarda doğru karar verme noktasında büyük bir fark oluşturan önemli bir çalışmadır. Sosyal medya içeriği mesaj, resim ve videolardan oluşur. Bazı durumlarda doğal afetin neden olduğu hasarı sadece metinden anlamak analiz açısından yeterli değildir, afetin etkisi görsel verilerle daha iyi anlaşılmaktadır. Sosyal medya platformlarından metin verisetleri araştırmacılar tarafında yaygın olarak kullanılmakta ve sınırlı sayıda çalışma görüntüler gibi farklı içeriklerin kullanımına odaklanmıştır. Bunun nedeni, afetlerle ilgili etiketli görüntü veri kümelerinin sayısının çok sınırlı olmasıdır.

Bu nedele bu tezde çok modlu bir Türkçe metin ve görüntü veri kümesi sunarak bu sınırlamayı ele almayı hedefliyoruz. Çalışmamızda, geç füzyon tekniği ile multimodal sınıflandırma çalışmaları yapılmıştır. Ayrıca, multimodal sınıflandırma elde etmek için; metni sınıflandırmak amacıyla; önceden eğitilmiş LSTM modeli kullanılırken, görsel içerik için önceden eğitilmiş bir CNN modeli kullanıyoruz. Genel olarak, çok modlu öğrenme mimarisinde her iki verinin birleştirilmesi ile %91,87 doğruluk sağlandı.

Anahtar Kelimeler: Çok modlu sınıflandırıcı, afet yönetimi, tweet metin sınıflandırma, tweet görüntü sınıflandırma, Türk Dili

To My Family

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisor Asst. Prof. Dr. Saed Alqaraleh for his approach and guidance throughout master process. I would also like to thank my jury members Asst. Prof. Dr. Bülent Haznedar and Asst. Prof. Dr. Muhammed Madi for their valuable participants and support.

This work is dedicated to my beloved daughters Ayşe and Ahsen, son Sirac, who have made me stronger and more fulfilled.

A special thanks to my mother, my father, my husband, and my siblings. I believe I achieve all of my goals with their support.

TABLE OF CONTENTS

Sayfa No.

LIST OF TABLES

Sayfa No.

LIST OF FIGURES

Sayfa No.

LIST OF ABBREVIATIONS

AdaBoost	:	AdaBoost Classifier
ANN	:	Artificial Neural Networks
BoW	:	Bag-of-Words
CBOW	:	Continuous Bag of Words
CMS	:	Crisis Management System
CNN	:	Convolutional Neural Network
DBM	:	Deep Boltzmann Machine
DL	:	Deep Learning
GBC	:	Gradient Boosting Classifier
GPU	:	Graphics processing unit
KNN	:	K-nearest neighbor
SGD	:	Stochastic Gradient Descent
LSTM	:	Long Short-Term Memory
ML	:	Machine Learning
MMDL	:	Multimodal Deep Learning
NB	:	Naïve Bayes
NLP	:	Natural Language Processing
RELU	:	Rectified Logical Unit
RF	:	Random Forest
SVM	:	Support Vector Machine
TF	:	Term Frequency
TF-IDF	:	Term Frequency-Inverse Document Frequency
Acc	:	Accuracy

CHAPTER I
INTRODUCTION

An extremely large number of images and texts are uploaded to social media platforms all over the world that are captured during most of the events occurring around us. This large-scale data shared on social media can be classified by visual recognition and textual understanding. In addition to the need to effectively extracting semantic content from multimodal data that has a complex structure, it is also necessary to carefully examine the nature of the data to be able to process it and get its information. Using social media data for social benefit is an important study, especially in the event of a natural disaster. Since natural disasters are time-critical situations, the effective classification of data published on social networks is extremely useful for humanitarian organizations to make plans and correct decisions on time. It is worth mentioning that sometimes people are sharing some information that is totally irrelevant to disasters with disasters hashtags some they ensure that more readers are seeing their tweets.

In this thesis, a multi-modal classifier application is carried out. First, Turkish text and visual files related to natural disasters are collected and a new annotated dataset is created. Secondly, experiments will be carried out on the introduced dataset to produce an efficient classifier. For this purpose, a general introduction is presented in the following section of this chapter.

1.1. Background

Nowadays, Twitter is one of the most popular microblogging that allows people to share what they do, feel, see instantly. Hence, it is a highly valuable source of information and for this purpose, it has gained importance for researchers. In this study, a multimodal classification system that processes both text and image data shared on Twitter before, during, and after natural disasters is presented.

Due to advances in deep learning, the performance of image classification methods has increased significantly in recent years. However, interest in multi-modal deep learning for image data and text data has increased. It demonstrates that deep representations of image data

and text data can be transferred to a new field by performing common deep learning representations for different data types.

1.2. Problem Statement

Disasters are known as events that have negative consequences on people, the environment and societies due to the life losses and the damages that can occur due to the disasters. Recently, messages and photographs are extremely used to describe the situations of people and the environment during natural disasters such as earthquakes, floods, and fires. Social media platforms, where information and news are accessed and used in seconds via the Internet, are considered one of the most widely used tools for communication and its purposes. In this case, such a platform, which we use for a large period in day manner, seems quite useful when in terms of the ease of accessing and availability of useful information. However, in cases of misuse, it creates a chaotic environment for communities and causes various harms. In other words, this misused platform may have the quality to have negative consequences. Hence, an automated system that can find out the most useful and relevant information before, during, or after disasters is extremely necessary. While many studies work on introducing efficient systems that automatically classify English data on social media, unfortunately, very few works have been done in this field related to the Turkish language.

Social media content consists of messages, images, and videos. Related to disasters, in most cases, understanding the damages only using text is hard, the effect of disaster is better understood with visual data. In recent years, the abundance of image and text data has become an essential factor to produce and develop deep learning systems of automated image and text classification.

Until now, most of the available disaster-related studies have been done using text only (Imran, 2015; Castillo, 2016; Kiatpanont, 2016; Pandey, 2016; Fersini, 2017; Nyugen, 2017; Taney, 2017; Alharbi, 2019; Jain, 2019; Wiegmann, 2020; Alam, 2021), and a few research works that deals with social media image data (Alam, 2017; Sakai, 2017; Hassan, 2020). And as mentioned before, all existing studies conducted to support the Turkish language focus only on the text. In other words, there is no Turkish study in which text and images are classified

together. Hence, we believe that images shared on social media during the crisis also contain valuable information in terms of classification.

1.3. Research Objectives

This study primarily aims to create a Turkish language multimodal dataset and an efficient system that can efficiently determine disaster situations. For this purpose, the following parts have been carried out:

- Collecting a large number of tweets and their contents (text and image), and then manually annotating these tweets using the annotators into two groups, i.e., disaster and non-disaster.
- Comparison of multiple classification machine learning algorithms such as Nearest_Neighbors, Linear_SVM, Polynomial_SVM, RBF_SVM, Gaussian_Process, Gradient_Boosting, Decision_Tree, Extra_Trees, Random_Forest, Neural_Net, AdaBoost, Naive_Bayes, QuadraticDiscriminantAnalysis (QDA), Stochastic Gradient Descent (SGD) using the created dataset.
- Building an efficient multimodal classification of CNN and LSTM deep learning techniques combined with late fusion.
- Evaluation and measurement of classification performance of the developed system created using the created dataset.

The performance of the machine learning algorithms mentioned above and the deep learning technique are compared using various criteria. These criteria are:

- Investigation of the preprocessing steps vs. using row data on the overall classification performance.
- Investigation of the effect of changing the number of samples used for training and testing the system.
- Investigation of the performance of some state-of-the-art method features extraction algorithms for text.

The method proposed in the thesis uses images and Turkish texts posted on Twitter during natural disasters. The success of the thesis will result in a disaster classification system based on social multimodal data and make the classification result more effective and accurate.

1.4. Scope of Work

After natural disasters happen, urgent and on-site intervention is very important to save lives and property and to make sure that such a situation is under control. One path that can help is to collect instantaneous online information and make an accurate analysis. Hence, numerous tweets are sent on Twitter before, during, and after natural disasters. Among these tweets, it is essential to analyze data that is relevant to the disaster and also ignore the irrelevant ones. It is worth mentioning that text can not always describe the disaster, while the images can better illustrate the disaster. In this thesis, a multi-modal system is designed in which the content of tweets (text and image) published before, during, and after the disaster can be classified as relevant or irrelevant.

1.5. Methodology

In this thesis, we evaluate the proposed model on datasets of five different disasters by adopting both text and image features. Hence, we represented each tweet using both its text and images. Related to the text representation, word embedding methods are applied after the preprocessing steps. On the other hand, for visual representation, feature extraction and data augmentation are applied to get visual characteristics. Then, in multimodal representation, the late fusion technique is applied to classify text and image tweets as relevant to or irrelevant to disaster.

1.6. Thesis Outline

In the following chapter, a review of literature, and the studies of the disaster management and data classification methods on social media are provided. The third chapter describes the text and image-based classification models. In the fourth chapter, the multimodal presented in the thesis is explained. In the fifth chapter, the results of the experimental studies

are explained by providing information about the datasets created for the deep learning model. Finally, the results of the study are mentioned in the sixth chapter as a conclusion.

CHAPTER II
LITERATURE REVIEW

2.1. Introduction

Nowadays, emergency managers are trying to use all available channels such as phones, TV, and Internet(websites such as social media) to inform the public about important information. The contents of the Internet environment where emergency management information is shared often use technical language that is difficult for a person to read, interpret or understand, which reduces the effectiveness of the emergency platform. Information on emergency management is typically time-sensitive, subject to constant change, and critical to society's readiness to respond to emergencies and disasters. In this context, as social media popularity increases, it is frequently used in a crisis. For example, when we look at the reactions given to the posts of the President and the members of the Council of Ministers during the July 15 coup attempt in Turkey, it is seen that the posts are shared an average of 300 times a day, retweeted 231,247 times, favorited 10,522 times and received 55,159 responses. (Demirtaş, Z.G, 2017). For this reason, it is important to use such sources to build an automated multimodel system that can classify and analyze whether the data shared on social media is related to disasters or not.

The concept of social media was first used by Barnes in 1954. Social media data can hold about zettabytes worth of data per day. These data are only meaningful when analyzed. As (Bontcheva, 2014) mentioned that the use of semantic technologies for smart information access and data classification in social networks is a very interesting and essential research area. In general, social media can be blogs or microblogs. Blogs allow unlimited content, while Microbloggers allow up to 140 characters of user-generated content such as single sentences, individual images, or video links. Because of this abbreviated structure, microblog posts are often represented using sentence sections, and abbreviations. Twitter is the most popular microblog, where approximately on average 309 new users are joining every minute. (Data Never Sleeps,2021) Twitter uses additional categorical tools called hashtags to categorize and rank posted content. These hashtags can be formed with letters, characters, and numbers following the hash ("#") and referencing an issue, event, or condition. These hashtags can be created by users.

2.2. Social Media and Twitter

Social media is an effective platform among the most preferred tools that provide communication and information flow. Social media has expanded the options for sharing information, changing the way people communicate in the event of a disaster. (Al-Akkad and Zimmerman, 2012; Palen and Hughes, 2018; Reuters, 2018). Especially in disaster and crises, people try to communicate more. For example, after a disaster, people want to know that their relatives are safe and at the same time, they want to get reliable information about the event and the post-disaster crises. In addition to pre-disaster awareness activities, social media is a quick way to access information during a disaster. In this process, it should be noted that the abuse of social media tools may increase crises. To prevent this situation, those who will provide information after the disaster should explain the information that needs to be reliable, clear, and accurate (Çetinkaya, 2013). In the event of a post-disaster crisis, information should be provided as often as possible, and the public should be informed about the situation of the disaster, thus preventing rumors and preventing the situation from turning into a secondary crisis. This helps to gain public trust and manage the process in a better way. Studies on the relationship between social media and crisis management have just begun. Therefore, studies on this subject are limited, but the crisis in social media is similar to the traditional crisis. For this reason, the crisis management of social media depends on the principles of classical crisis management. However, (Diyadin & Özdil, 2017) suggest that these methods and mechanisms should be further improved for social media.

For example, during 2011 a magnitude 5.8 earthquake that occurred in the United States, authorities have contacted the public via Twitter to report the disaster damage in their regions and to inform the public what to do. In this earthquake, where calls could not be made due to equipment disruptions, Twitter was used to reach people's relatives and get information from public institutions (Soydan and Alparslan, 2014). Another example is the Van Earthquake that took place on October 23, 2011, in Turkey. In this event, people in Van and surrounding places, have organized aid, support campaigns, and rescue activities using social media. The fact that the injured are under the rubble in some regions shows how important social media is in the context of correct use. It should be considered that platforms can be used effectively and also help in crisis management.

However, due to the extremely huge amount of shared multimodal data, it is important to remove both redundant and irrelevant information to assist decision-makers in making the most suitable actions during such events. Researchers are interested in big data in social media, and it is widely used to analyze the textual content only of social media. With the increase in Internet speed, the analysis of image information is also becoming widespread. While most current event detection approaches generally focus on text content, few researchers explore multimodal information among heterogeneous data. Studies in this field are mainly classified as unimodal. In other words, in current studies, image and text data are evaluated separately and there is no multimodal study for supporting and processing the Turkish language in the literature.

In this study, we use the Twitter social media platform due to many factors, such as Twitter is considered as one of the most widely used social media networks with approximately 310 million monthly active users, 83% of which are on mobile devices that makes providing first-hand information during a disaster more frequent. Also, Twitter users tweet with many different data modes, including text and images. To integrate these different data types and extract relevant information, we built a multimodal neural network combining Turkish texts and images. Overall, in this thesis, an efficient multimodal approach was developed to classify images and text using deep learning algorithms.

According to (Our World in Data, 2021), the earth's population is around 7.7 billion and at least 3.5 billion users are online. Also, it has been reported that Twitter gains 319 new users every second, and 473,400 tweets are sent every second. Table 2.1 summarizes what happens on the Web each one minute. Also, the number of social media users in Turkey is equivalent to 70.8% of the population according to datareportal's January 2021 report (Datareportal, 2021) Twitter, one of the most used social media platforms in Turkey, is used by 72.5% of the Internet users. The main reason why Twitter is so heavily used is that textual and visual information can be transmitted so fast.

Table 2.1. A minute on the Internet in 2021 (Our World in Data, 2021)

Social Media Platform	Every 1 Minute of the Day
Twitter	Users post 575k tweets
Tiktok	Users watch 167M videos
Instagram	Users share 65k Photos
Facebook	Users share 240k Photos
Youtube	Users stream 694k Hours
Netflix	Users stream 452k Hours
Google	Conducts 5.7M searches
Teams	Connects 100k users
Zoom	Hosts 856 Minutes of Webinars
Snapchat	Users send 2M Snapchats
Amazon	Customers spend $283k

In the following, the most recent and related works and literature to this study and the methods used to classify disaster-related social media data are summarized. In (Rudra, 2018), authors proposed distinctive features to classify tweets as contextual and non-contextual. After classifying the tweets, they used situational tweets summarization techniques to convey awareness to government agencies. Then, their proposed system was experimented with different disaster datasets such as the Uttarakhand floods and the Nepal earthquakes, using tweets in two different languages, English and Hindi. They developed a domain-independent classifier that performs better than the domain-dependent technique in the case of both English and Hindi tweets.

(Caragea C, 2014) also categorizes building damage by using text-only resources from affected individuals to provide easy access to humanitarian organizations and disaster-affected people. Similarly, (Rudra K, 2017) the authors proposed an automated method to classify tweets about Ebola and MERS. In (Madichetty, 2019), the authors developed a method to classify tweets about damage detection, combining statistical and illuminating features. They used Random Forest (Breiman, 2001) and AdaBoost (Freund, 1996) classifiers. However, these mentioned articles talked about different types of useful information and only textual features were explored. Detecting informative tweets during a crisis is crucial. (Nguyen, 2017) also proposed a technique using visual content during a disaster, but did not explore any textual features. There is no study in the literature to classify Turkish informative tweets with a multimodal classifier that uses both picture and text content. This thesis proposes a model that

uses Turkish tweet text and image content to classify relevant or irrelevant tweets during a natural disaster.

(Castillo, 2016) suggests in his study that it is very necessary to use social media microblogging platforms such as Twitter for rapid and rapid damage assessment during disasters. They have obtained 80% accuracy by classifying 52 million tweets using Naive Bayes, Support Vector Machine, and Random Forest methods. It is difficult to obtain useful tweets by filtering tweets to get information such as how many people were injured or killed, how many buildings collapsed, but also urgent action may be required by various organizations such as what medical supplies the survivors may need. Therefore, tweets during a natural disaster can be classified as informative and non-informative. Informative tweets (Houston, 2015) are written and visual data that provide information about disaster-affected people, the needs of the injured, and the availability of necessary medical supplies.

The author (Alqaraleh, 2021) aims to create a system that can detect crises that require assistance by making an effective classification for Turkish tweets and this study only classifies text using KNN, SVM, and CNN algorithms and has obtained 94.88 as accuracy. Another Turkish study was presented in (Ayata, 2017) where word vector representation was used instead of rule and dictionary-based approaches. Decision Support Machine and Random Forest classification are used, and a study is presented in which four different datasets are classified for sentiment analysis in Turkish.

In the study (Alam, 2018), a graphical deep learning model was developed with an inductive semi-controlled technique using low-label data in two different categories of tweet datasets to respond instantly to the crisis due to the lack of labeled data. In the study (Alrashdi, 2018), the performance improvement of the model was made by using The CrisisNLP tweet database, BiLSTM and CNN network components, and the Glove word embedding model for crisis management classification. The article (Alam, 2018) introduced an image processing pipeline to collect images, remove duplicates, filter out what seems irrelevant, and classify according to the degree of damage to extract meaningful data from social media visual content during a natural disaster.

Information fusion can occur in two stages, early and late fusion levels. It has been proven in many studies that the late fusion approach is more effective (Ces G. M. Snoek, 2005). Most importantly, since the modules that extract semantic information from the image, sound, and text in the project work independently of each other, it is more appropriate for our fusion approach to take place at the late fusion level in our project. (Rizk, 2019) using the SUN dataset, the earthquakes in Nepal, Chile, and Japan and the tweets of the flood disaster in Kenya were classified using SVM and ANN and an accuracy of 92.43% was obtained. In (Jony, 2021), the author classifies 6000 Flick images in two categories, flood or non-flood, using the MediaEval 2017 dataset, with the CNN algorithm. (Zou, 2021) image and text classification was made using the CrisisMMD dataset obtained from the events that took place in 2017. Until now, studies in which text and image data are used together in this way have been done in English.

Unfortunately, until the time of this study, no multimodel crisis management system study classified text and images made for the Turkish language. However, with this thesis, we consider the crisis management system with two different classifications and use multimodal deep learning techniques. In our study, both text and image-based Turkish data were created, and multimodal classification studies were carried out with the late fusion technique.

CHAPTER III
CLASSIFICATION METHODS

Within the scope of this thesis, a system was developed in which automatic semantic information was extracted from Turkish tweets using image and text data (multi-modal), stored in appropriate formats, and then effectively classified. Within the scope of the thesis, semantic information is extracted from each of the image and text data. When textual and visual methods are combined, it seems better to use these two contents simultaneously (Bouslimi et al., 2017). In short, this process, i.e., multi-modal classification, has been one of the important activities of the thesis. The following sections present the problems, mechanisms, and steps of deep learning classification models when processing both text and image.

3.1. Text and Image Classification Problems

The field of natural language, includes many computational techniques for representing and automatic analysis of human language. It is hard and difficult to perform text classification using computers and machines. (Cambria and White, 2014) described the anticipated development of NLP research with three different curves to overcome difficulties in this area: syntactic, semantics, and pragmatics. Most of the approaches available today still use the syntactic representation of the text. The second trend, semantics, study the meaning of the text by concentrating on its content. While texts are represented by the bag of words model in syntactic tendency, i.e., the text is represented by the bag of concepts model in semantic tendency, where the bag of concepts is based on models that represent the meanings of words in a vector space. In the bag of narratives model, which shows the pragmatic tendency, it is not foreseen that a more detailed text comprehension and calculation level will be achieved by showing each piece of text with smaller and interconnected sections. Many factors complicate the text classification process. One of these factors is that there are hundreds of different natural languages, each with different structures and rules. Another difficulty can be ambiguous, as the meaning of words changes depending on the context.

On the other hand, when we consider a series of images tagged with one category, these categories are expected to be estimated for a new set of test images and the accuracy of the predictions is measured. In doing so, many problems such as viewpoint variation, scale

variation, image distortion are arisen. In the following, we summarized the main problems related to process text and images.

3.1.1. Selecting the Appropriate Feature Extraction Method

It is one of the most significant parts of machine learning systems. Because the correct operation of the system varies depending on the selection of the right features and the number of features. However, the property inference process is domain-dependent, which means that the properties selected for the facial recognition system and the properties selected for the fingerprint recognition system will be different. Likewise, the feature selection of the system developed for an English text classification task will be different from that for Turkish. If machine learning features can be learned automatically, the entire learning process can be performed automatically, making it easier and many problems can be solved. At this point, deep learning has provided an easy way to provide automated feature learning.

3.1.2. Noise

In a facial recognition process, faces in images can be of different sizes, colors, angles, lights, or low resolution. In a text classification process, texts from different site sources may contain misspellings, punctuation, non-widely used abbreviations, missing sentences, and texts. Figure 3.1. illustrates the impact of noise on the classification of a simple, linearly separable, binary classification dataset.

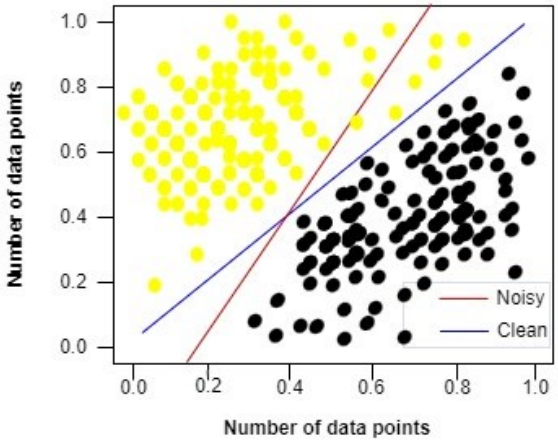

Figure 3.1. Examples of comparative scatter plots in two types of data (noisy and clean data)

3.1.3. Underfitting

If a nonlinear function is modeled with a linear model, the model falls short of expectations and the bias error is high. When bias error is high, training data is learned incompletely. In short, this condition is called underfitting. A network with a high bias error generates prediction outputs on the training set, a high error rate, and the model does not learn the data well. To reduce bias error, it is necessary to modify the network architecture by adding new layers and neurons. CNN and RNN methods can be a good solution to this situation.

3.1.4. Overfitting

Variance error is an error that occurs because of randomness in the training data. When the training distribution is improved so that the model can no longer be generalized, a high variance error occurs. This condition is called overfitting. A network with a high variance error has learned the training data very well and has faced the problem of memorizing the data. At the point where the validation error is separated from the training error and increases, the problem of overlearning begins to occur. Increasing the number of data to reduce variance error, applying dropout and regularization techniques can be the solution. In a facial recognition or text classification process, each training sample can be classified very well by increasing the model complexity. However, the model's performance on test data, i.e. previously unprecedented data, may be poor. If a model is performing very well in training samples and performing poorly in test data, it has over-learned. The error of the model trained on the training data does not provide the correct information, so it is necessary to look at the performance, generalization error, on the test samples of the trained model. Figure 3.2 shows the concepts of underfitting, balanced, and overfit. The graph on the left is where we can predict that the line does not cover all the points on the graph to explain the variance, causing the data to be missing, thus having a high bias error. In contrast, the graph on the right may look like a useful graph that covers all points; however, the line projected in the graph incorporates noise and random fluctuations in the training data into its learned relation. Such models most likely predict poor outcomes due to complexity. The model has a high variance error. Finally, the middle chart shows the proper fit as it has a pretty good, predicted line. The line covers most of the points on the chart and we can conclude that there is a balance between bias and variance.

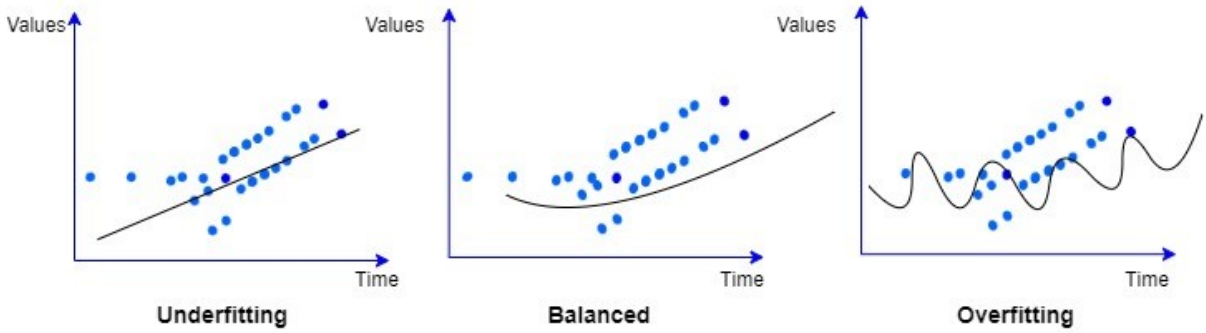

Figure 3.2. The concepts of underfitting, balanced, and overfitting

3.1.5. Selection of Learning Algorithm

The model used for a particular job may not work correctly and alternative algorithms may be needed. The researcher needs to test many models and develop or use the algorithm that suits the need to select a good learning algorithm for the need of the problem. In addition, having preliminary knowledge of descriptive properties for a recognition or classification problem is important in creating more original models for separating different types of classes.

3.1.6 Missing Values

The missing features are mainly due to a lack of data or the use of the opt-out option. For example, faces in some races may not be included in the training set due to difficulty in recognizing, or they may be excluded from the class category because the news classification process has little data in the culture class.

The following section can be divided into two subsections. The first subsection describes the classification process with text, and the second subsection describes the classification process with images.

3.2. Text-Based Classification

It has emerged as an effective solution to text analytics problems to use deep learning. Recently it has improved the performance in many fields related such as classification, clustering of texts, document summarization, customer relationship management, web mining, emotion analysis, etc.

Text classification is the determination of whether each document (text) belongs to predefined classes. In other words, a logical value in the form of true or false must be produced for each input. The methods used in the text classification process are explained in the following section.

3.2.1. Preprocessing and Indexing

Textual data has been pre-processed because they contain terms such as spaces, punctuation, repetitive words, etc. Turkish-specific characters " ç, ğ , ı, ö, ş, ü" are converted to Unicode characters. One of the pre-procedures performed is "stop words", which is the process of removing words such as unnecessary conjunction and preposition from the text. Due to the sensitivity of uppercase letters, all words in the text have been converted to lowercase letters. Unnecessary characters have also been cleared during the pretreating process. A sample procedure is also shown in Figure 3.3. for the operations performed.

Tokenize: It also removes punctuation from the text by separating the string into tokens.

Stop word filtering: Eliminates common words.

Stemming filtering: Reduces each word to its body by removing prefixes or semed attachments.

Indexing: After filtering and generating data, it is indexed using TF-IDF, a weighting scheme that scans features in tweets, depending on how often the word occurs in a single tweet compared to how often it occurs in other tweets. Then, by measuring the weight of each keyword, the relevant event can be decided.

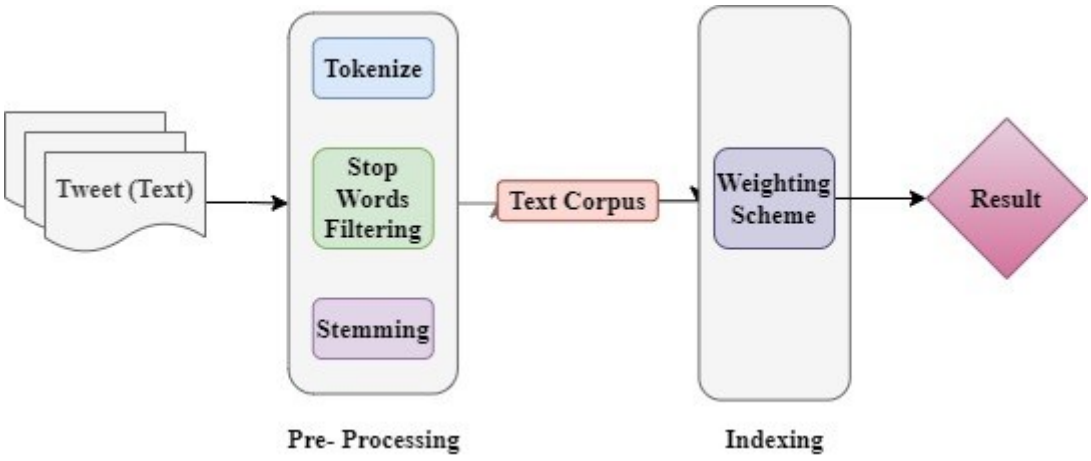

Figure 3.3. The block diagram of text data preprocessing and indexing.

3.2.2. Vector Representation of Texts

As with traditional machine learning models in deep learning models, the texts given before they can be trained must be converted to numerical representation. Converting text to numeric representation is called vectorization, which can be done in several different ways:

- Convert text to words and show each word as a vector
- Convert text to characters and show each character as a vector
- Create n-grams of words and represent them with vectors

The separation of a given sentence into characters or words is called a unit (token). After text data is divided into small volumes, each unit is mapped to a vector. There are two basic approaches for matching units with vectors: 1-hot encoding and word embedding. In this thesis, the BERT algorithm was preferred to create word embeddings. BERT offers powerful architecture. For this reason, the bert-base-turkish-uncased (Loodos,2020) model, which is a pre-trained BERT model using MLM technique with a 200GB data set obtained by the Loodos team from sources such as e-books, news articles, online blog posts, and Wikipedia, was preferred for the Turkish language.

Although Word2vec is a very successful technique, there are some difficulties; although the word2vec model is easy to develop for a new data set to be created, debugging difficulty is one of the most important problems. In the case where a word has more than one meaning, the

word representation will average these meanings in vector space. The uncertainty that will occur in this case is another challenge of the word2vec model.

3.2.3. Word Embeddings

3.2.3.1 BERT (Bidirectional Encoder Demonstration) Transformers

The algorithm started to be used in 2018 with other algorithms of Google in NLP to better understand the queries and to provide more accurate results by evaluating the words in the queries one by one. Using artificial intelligence and machine learning technologies together, the BERT algorithm can be used to find high-quality language from your text data. In this algorithm, instead of evaluating the words in the queries one by one, they will begin to evaluate them together with the words before and after them or similar and synonyms. BERT means that with the Algorithm update, the conjunctions and prepositions used in the queries will understand much better what they add to the query. Regardless of what appears in the Word2vec model in the context required for it, BERT is transmitted dynamically informed word representations with words around them. In this thesis, BERT was used to create pre-trained word representations in twitter texts.

3.2.3.2.ELMO

Elmo word vectors are used in two bidirectional language model layers. There are two transitions in its layer; forward transition and backward transition. Advanced transition; It contains information about a specific word and previous words. The passback contains knowledge about the word and the scope after it. This information creates intermediate word vectors in forwarding and backward transition. These word vectors feed into the next layer of the bidirectional language model. As the final indication, Elmo is the weighted sum of the raw word vectors and two intermediate word vectors. The same word can have different word vectors in different contexts if the whole sentence with ELMO word vector vectors is the process. Thus, it provides an advantage over traditional word vectors such as Word2vec and Glove, which have a single numerical representation regardless of where the words occur and the fields that have different meanings.

3.2.3.3.XLNet

XLnet is the newest and largest model for NLP. XLNet is based on the recursive transformer architecture and is a language model that gives the common persistence of a set of words. The purpose of being used in modeling computes word vector calculations based on permutations of all word vectors in a sentence, as expected by the source stem and its left. While doing this, BERT avoids the assumption of independence and fine-tuning.

3.2.3.4.Fasttext

The advantage of FastText representations over word2vec representations is that they take into account the inner structure of words when learning word vectors. This feature is useful for morphologically rich languages such as Turkish. When training in the Word2vec algorithm, every word is considered atomic, i.e. unique, but with FastText, each word is grouped according to n-gram characters.

3.2.4. LSTM Text Classification Model

Long Short-Term Memory Networks (LSTM) Model Memorization-overlap and reset gradient (vanishing gradient) are two main problems in deep neural networks applications. The dropout method is used to avoid the problem of memorization due to the raised parameters' numbers in the hidden layer. (N. Srivastava, 2014).

The RNN method, which is widely used in NLP, can predict the latter word from the former words given in a content. Like traditional neural networks, RNN uses a reverse propagation algorithm. During this backward spread, gradients prone to zero, called the reset gradient problem. LSTM architecture is introduced to solve this problem (S. Hochreiter, 1997).

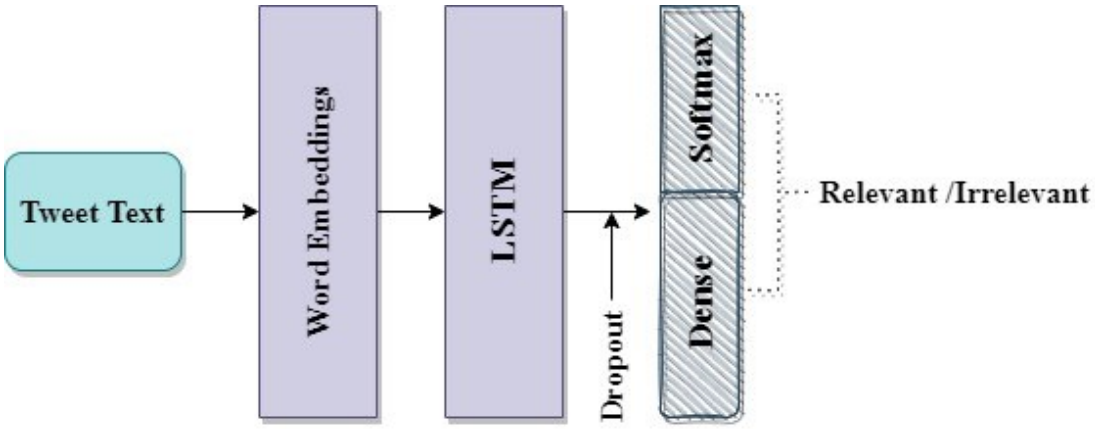

Figure 3.4. LSTM network model architecture.

Figure 3.4 shows the architecture of the single-layer LSTM utilized in this thesis, the LSTM layer is already trying to determine whether a tweet given by training an RNN is related to a natural disaster. Depending on whether they are relevant or irrelevant, the lengths of comments of different lengths labeled 1 and 0 respectively are limited to 100 words. Longer comments are cut according to the maximum sentence length, while shorter comments are filled with zero padding, i.e. the required amount of zeros. A dropout was used after this layer to prevent overfitting.

The following section describes the image classification process. Image classification is the function of extracting the probabilities of which image class the image belongs to or the classes that best describe it for the image after the image is taken as input.

3.3. Image-Based Classification

When people look at an image, they can effortlessly differentiate between the color, size, and similar types of items. When computers detect the same images, they create a matrix. A matrix value for each pixel in the image varies according to that pixel. Convolutional neural networks process these matrices in a large number of hidden layers, detecting their different properties so that they can easily distinguish objects. Today, the rapidly evolving method of image processing is a technology used in many fields to perform operations such as object recognition and classification by converting an image into a numeric form. Image classification is the process of separating pixels in an image under different class labels using different interpretation techniques. (Figure 3.5) The image classification includes tags that are classified

as outputs. In this thesis, class labels on the image are used in supervised classification, which is predetermined by the user. Support vector machines, Bayes classification algorithm, K-nearest neighbor algorithm, decision trees, and convolutional neural networks are used in disaster classification.

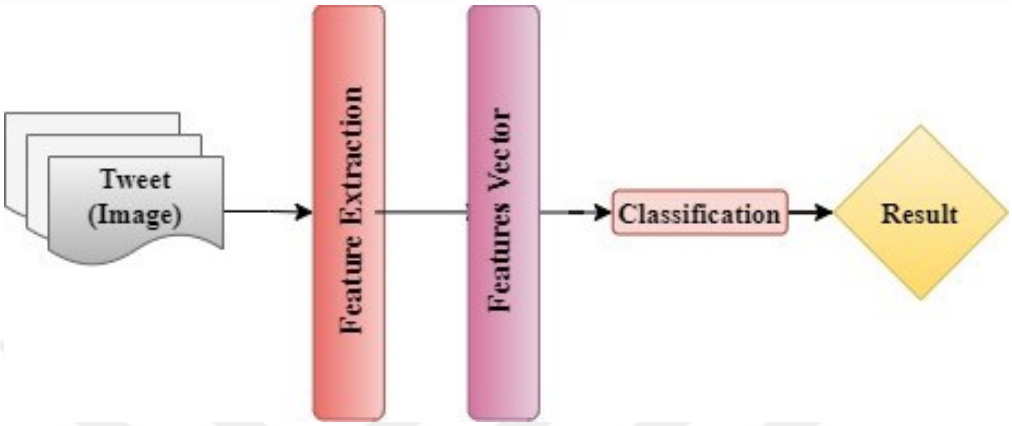

Figure 3.5. The block diagram of image classification

3.3.1. Convolutional Neural Networks

Convolutional Neural Networks (CNN) play an important role in the creation of classification models based on image recognition and natural language processing. Convolutional Neural Networks, a category of deep learning neural networks, has performed superhumanly in many other areas of image recognition, object recognition, automatic video classification, and computer vision. Convolution Neural Networks consists of a combination of popular convolution processes and neural networks. Convolutional Neural Networks (CNN or ConvNet) is a type of multilayer sensors (MLP). The cells in the visual center are divided into sub-regions to cover the entire image. Simple cells concentrate on edge-like properties, while complex cells concentrate on the entire image with larger receivers. CNN, an advanced neural network, was inspired by the animals' visual center. Although it was first reported in 1998 that Yann LeCun published an article, Lecun, Bottou, 1998, it was made known worldwide in 2012. The AlexNet model, designed with deep learning architecture, won the ImageNet competition held that year. The study was published with the article "ImageNet Classification with Deep Convolutional Networks" (Krizhevsky, Sutskever, et al. 2012) and quoted 21944 as of July 2020. With this architecture, the computed object identification error rate was degraded from 26.2% to 15.4%.

Convolutional neural networks are the most trendy neural network model used for image classification, as they greatly improve learning time with fewer parameters and reduce the amount of data needed to train the model. Instead of a network fully connected to each pixel, CNN weighs enough to get a small portion of the image. The main layers of CNN are described in the following.

3.3.1.1. Input Layer

This layer constitutes the first layer of CNN. Data is given raw to the network. For the success of the model to be designed, it is important to select the correct input image size of the data in that layer. When the input image size is selected high, the requirement for memory increases, and the training and test time can be extended. If the input image size is selected as low, the performance of the network may be low while the memory requirement and training time is reduced. When performing image analysis, an appropriate input image size is needed to choose for network success in terms of both network depth and hardware cost.

3.3.1.2. Convolution Layer

In evolution, which is a customized linear process, the main purpose is to extract properties or information from an image or text. These networks are simply networks that perform convolution rather than matrix product at least one layer. (Ethem, 2018)

Discrete time convolution is expressed in Equation (3.1)

$$s(t) = (x * w)(t) = \sum_{\alpha=-\infty}^{\infty} x(\alpha)w(t-\alpha) \tag{3.1}$$

In Equation (3.1), filter w, input x, time t, and s are expressed as results. When a two-dimensional input, such as an image, is used as an input, the convolution process is expressed by equation (3.2). (Ethem, 2018)

$$s(i,j) = (I * K)(i,j) = \sum_{m}\sum_{n} I(i,j)K(i-m, j-n) \tag{3.2}$$

In equation (3.2), the terms i and j refer to the locations of the new matrix to be obtained as a result of the convolution process. In many cases, the center of the filter is positioned to be in its origin.

An image or a text is considered a matrix of values, and the purpose of the convolution process is to scan this matrix with specific filters to extract descriptive properties for images and texts. Filters are an integral part of layered architecture. Filters can be of different sizes such as 2x2, 3x3, 5x5. Filters generate output data by implementing the convolution process to images from the previous layer. As a result of this convolution process, the activation map is created. Activation maps are regions where characteristics specific to each filter are found. During the training of CNN's, the coefficients of these filters modify with each learning loop in the training set. Thus, the network identifies which regions of the data are significant for determining properties. For the convolution process, 3x3 filterscrolls through the input image to the right or left by swiping the specific step. During this entangling, when the matrix boundary is reached, one step down and continues again. This entangling is done on the entire image matrix. Filter coefficients are multiplied by values in each color channel and their sum is taken. After this operation is performed on all three channels, the sum of the three creates the activation map. Filter coefficients applied to each color channel matrix may be different. Changes in these filter coefficients are made by designers to suit their models. The values on the activation map are normalized to provide the same intensity range between the input and output size. For this normalization, the calculated values for each color channel are divided by the sum of the filter coefficients.

3.3.1.3. Flattened Linear Unit Layer

This layer is used after the convolution layers and is known as the rectifier unit, which is most activated for the output of CNN neurons. The ReLu function takes 0 for negative output as given in Equation 3.3, while it takes its value for positive entries x

$$f(x) = \max(0, x) \tag{3.3}$$

This layer is also known as the activation layer. The impact it has on input data is that it lowers negative values to zero. The network is linear because certain mathematical operations were carried out on the convolution layer utilized before this layer. This layer is implemented to put this deep network into a nonlinear structure. With the use of this layer, the network learns faster.

3.3.1.4. Pooling Layer

Pooling is usually positioned after the ReLu layer. Its primary purpose is to reduce the input size for width and height for the next condensation layer. This does not interfere with the depth size of the data. The process performed on this layer is also called "DownSampling". The decrease in size because of this layer leads to the loss of information. It is useful for two reasons for such a lost network. The first creates less transactional load for the following network layers. The second avoids the system from memorization. Like the convolution process performed in the first step, certain filters are identified in the pooling layer. These filters are routed around the image according to a particular step-by-step value and placed in the output matrix by taking the maximum values (maximum pooling) or the average of the values (average pooling) of the pixels in the image. Usually, maximum pooling is favored since it accomplishes better. Pooling is performed for all images because of the convolution layer.

3.3.1.5. Fully Connected Layer

In the CNN architecture comes the full connected layer after the convolution, ReLu, and pooling layer. This layer connects all areas of the previous layer. The entire matrix is given a single class vector with a size of 1x1x10. The number of neurons in this layer is equal to the number of classes, and the layer output gives the class score, that is, the probability that the given input owns a class. As a result, an input passes through a series of layers, turning into an estimated class score in the output layer. The parameter or weight requirement of each layer is different, and these network parameters are learned by gradient reduction-based algorithms in the process of propagation.

3.3.1.6. Dropout Layer

CNN sometimes memorizes the network because it's trained with big data. This layer is used to avoid the network from memorization (Srivastava, Hinton, et al. 2014). The basic logic implemented on this layer is the unload of some nodes of the network.

3.3.1.7. Classification Layer

This layer comes after the fully connected layer. Classification is carried out in this layer of deep learning models. The output value of this layer is equal to the number of objects to classify. For example, if 10 different objects are to be classified, the classification layer output value must be 10. If the output value is selected as 4096 on the fully connected layer, a weight matrix of 4096x10 is obtained for the classification layer according to this output value. Different classifiers are used in this layer. Mostly softmax classifier is favored due to its achievement. In the classification, 15 different objects produce outputs of a certain value in the range of 0-1. The output, which produces close to 1 result, is confirmed to be the object that the network presupposes.

All the layers explained above include hyperparameters, such as the number or shape of the filters of each Convolutional layer.

The best invention in the competition, held in 2012, was the AlexNet architecture developed by Alex Krizhevsky at the University of Toronto. In 2014, two architectures came to the fore in the competition. The first is the VGG architecture introduced by the visual geometry group at Oxford University. There is two edition of this architecture, VGG-16, and VGG-19, with 16 and 19 layers, respectively. The second is the GoogLeNet architecture, known as Inception. In 2015, it won the network architecture competition, called ResNet, presented by the Microsoft research group. These popular CNN architectures are described below.

A) AlexNet Architecture

The first popular use of deep learning in computer vision started with the AlexNet model. This 8-layer architectural array, aims to classify images in the ImageNet database with

10 million images and 1000 different image categories. The AlexNet model has all pooling layers in size 3x3, 4096 neurons in The Full Link-1 and Full Link-2 layers, and 1000 neurons in the last layer, Full Link-3. The latest layer, therefore, represents 1000 classes. AlexNet is also a pioneer of the relu activation function and dropout technique in deep neural networks.

B) VGG Architecture

Designed by Simonyan and his counterparts in 2014, this architecture is based on the insight that deeper networks are better networks. Despite it providing higher accuracy performance than AlexNet, it has a lot of parameters (about 140 million) and needs a lot of memory usage. In other words, smaller filters are used according to AlexNet. This architecture utilizes fixed 3x3D filters with a variable number of 64, 128, 256 filters in all convolution layers.

C) Inception Architecture

The biggest contribution of this architecture is to reduce the number of parameters to 5 million (approximately 12 times fewer parameters) compared to the AlexNet architecture, which has a total of 22 layers and a total number of parameters of 60 million. There are four versions: Inception-v1, Inception-v2, Inception-v3, and Inceptionv4.

D) ResNet Architecture

Researchers encountered a problem creating deep CNN architecture. Previous architectures have seen that as you add layers to deep learning models, performance increases up to a point, and at some point, there is a rapid decline. This problem, known as a reset gradient, was spread back during network training. The state-of-the-art ResNet architecture in image recognition, like previous architectures, is built on the idea that "the deeper the network, the greater the performance". However, with increased network depth, the reset gradient problem is also increasing because the gradient of each layer is calculated by the chaining rule compared to gradients from the previous layer.

3.3.2. CNN Hyperparameters

The parameters determined before the training process are known as hyper-parameters (Tran, 2020). Before applying any method to the dataset, it is necessary to select an appropriate set of hyper-parameters (Young, 2015). Hyper-parameters have a significant impact on the performance of the trained model. Different models can be created by specifying different sets of hyper-parameters. (Tran, 2020). Below are some hyper-parameters specific to the CNN deep learning algorithm.

3.3.2.1. Activation Functions

The activation function is used while obtaining net output results by processing the input data for nonlinear transformation operations in the multilayer neural network. Different activation functions have been developed for artificial neural networks to solve complex problems over time. Thus, the artificial neural network is used to collect the information coming to the nerve cells, and the selected activation function is used to transmit the output value in the neurons to the next layers.

3.3.2.1.1. RELU Function

The task of this function is that if the input value is below zero, the output is zero, and if the input value is above zero, the output creates a linear relationship by equalizing the input value. Compared to other activation functions, it produces fast results for computers with limited processing capacity.

3.3.2.1.2. Sigmoid Function

Used for nonlinear problems. Logistic and tangent functions are called sigmoid character functions. The logistic function applies to the input vector values separately and generates values between 0 and 1, while the tangent function maps the input value between -1 and +1.

3.3.2.1.3. Softmax Function

The softmax function generates predictions for each class and outputs the estimated class probabilities of the input data. The estimated class is determined by the highest softmax value. Maps the input data to probabilities with a sum of 1 between values 0 and 1 and maximizes probability.

3.3.2.1.4. Loss function

Loss functions are used to measure the performance of the model while training the artificial neural network models. Many loss functions calculate the difference between the target variable and the output.

3.3.2.1.5. Cross entropy

Cross entropy is expressed as a probability value whose output is between 0 and 1. The farther the predicted probability is from the original tag, the higher the value of the cross-entropy loss, the closer it is to zero. The cross-entropy function result is always positive. Cross entropy loss and mean squared error are the loss functions used to compare training data with real tags and determine the error amount.

3.3.2.1.6. Mean Square Error (MSE)

Mean Square Error is one of the most widely used loss functions. MSE, as its name suggests, is the sum of the squared distances of errors between our target variable and our predicted values.

3.3.2.2. Optimization Algorithms

In deep learning applications, the absolute minimum value of the error function must be found for the learning process to end correctly. This process is carried out using optimization algorithms. Optimization is the method used to make difference between the output value produced by the network and the actual value, that is, the error the smallest. There are 4

optimization methods commonly used to minimize error rates in machine learning. These are the ones that are going to SGD, Adagrad, RMSProp, and Adam methods.

3.3.2.2.1. SGD

In this algorithm, the training data is calculated with a specific training sample, not completely. In this way, memory-related problem occurrence is reduced.

3.3.2.2.2. ADAGRAD

In this algorithm, t different learning coefficients are used for each parameter at each step. The learning coefficient is not calculated manually and different updates are made for each parameter. Each parameter has its learning speed. In the education process, the learning coefficient gets smaller.

3.3.2.2.3 RMSPROP

This algorithm has been developed as an alternative to the extreme shrinkage of the Adagrad learning function. Instead of using all of the past education values like Adagrad, the exponentially weighted average of the values selected to a certain extent is used.

3.3.2.2.4. ADAM

This method uses weighted averages such as RMSPROP, and momentum changes are also cached. It has small memory requirements and is well suited for large amounts of data and parameter calculations.

CHAPTER IV
THE PROPOSED MULTIMODAL LEARNING APPROACH

Learning algorithms or models use different algorithms that correspond to produce sets of assumptions and representations for the input dataset. Different hyperparameters settings are made to improve learning result performance. A different approach is also carried out on different educational data. In addition, the most promising approach uses data from different methods and inputs such as processing both text and images of the input sample to provide efficient representations to the input sample, and this is known as multi-modal learning (Ethem, 2018).

The aim of this thesis is to fusion the information obtained from some data collected from Twitter where both image and text have been collected and a new dataset in the Turkish language related to natural disasters has been created. Then, using three evaluators, the collected samples were annotated as relevant to or irrelevant to natural disasters. Next, the created dataset was used to build automated and multimodal classifiers that support the Turkish language. In other words, social media content consists of messages, images, and videos. In some cases, understanding the damage caused by natural disasters only from text is not enough in terms of analysis, the effect of disaster is better understood using visual data. Text datasets from social media platforms are widely used by researchers, and a limited number of studies have focused on the use of other content such as images. This is because the number of tagged image datasets related to disasters is very limited. Therefore, in this thesis, we aim to address this limitation by presenting a Multimodal Turkish text and images dataset.

Recently, plenty of visual data, automated image classification tools have become important items of the task of deep learning architectures. Visual contents provide varied multimodal qualities, it offers a variety of sources of information that can be used to assist in classification. Also, in this study, we examine the effect of textual and visual methods on multimodal classifiers. We extract Turkish text and images from Twitter to process text and visual contents. We present a deep learning approach using LSTM for text, while the visual data(images) are processed using deep features extracted from a fully connected layer of Alexnet based CNN. Finally, we perform a late fusion of classification scores. Multimedia applications are very rich in content.

In general, as shown in the literature reviews, most of them existing studies are focusing and use a single data type, either text or image. The number of studies that use both image and text data is very limited. In this respect, it is evaluated that this study will make an important contribution to the literature. The results of scientific studies on the automatic extraction, storage, and querying of semantic contents of multimedia applications are not yet satisfactory enough. Therefore, more scientific research is needed on multimedia. In the coming period, the subject will continue to be a hot and essential research topic. We apply these approaches (CNN and LSTM workflows, respectively) in both visual data and textual data. Experimental results show that the LSTM workflow performs better in textual data, and the CNN approach is better at image modality. The relationship between text and visual methods is based on the underlying dataset and annotation. As shown in Figure 4.1, we verify the integration of visual and textual methods with multimodal techniques that perform better than single-mode baselines per dataset. Briefly, Figure 4.1, shows the workflow of the proposed multimode.

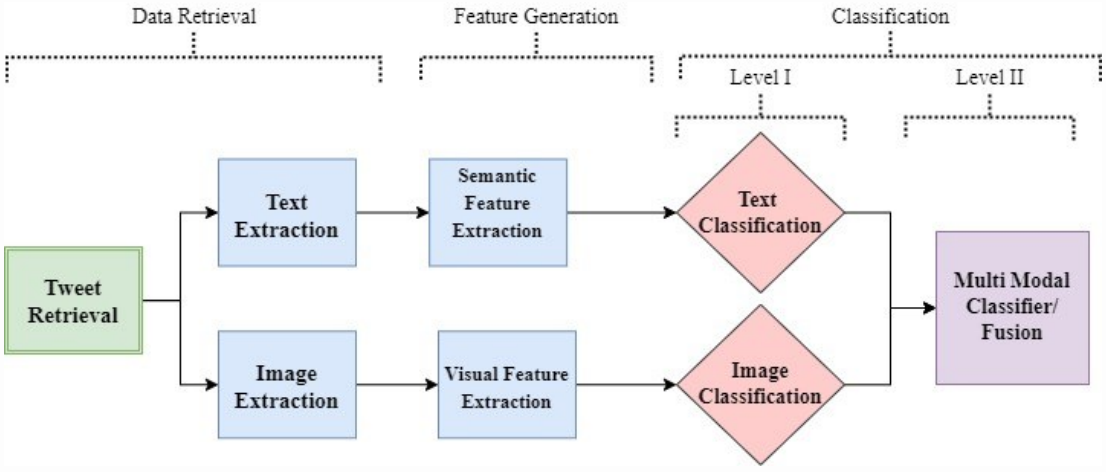

Figure 4.1. The workflow of the proposed multi-modal classification system.

In general, there are two basic approaches to automatically merging different models, in our case text and images, early fusion and late fusion (Elahi et al., 2010). Early fusion, working on combining the modes before classification (Meüller et al., 2010). For example, if you want to use If the properties are shown with vectors, a simple way of early fusion is to connect or combine vectors (Yu et al., 2016). On the other hand, late fusion works on finding the characteristics of each model separately and then combining the produced output (Wang

and Li, 2017). As shown in (Meüller et al., 2010), late fusion is usually and, in most cases, when applied to multimode information acquisition outperforms early fusion.

In more detail, early fusion (CONCAT) works on finding the best features from each multi-data (text and image), then each property is scaled to have zero average and standard deviation. Then, the features of both text and image are combined in a single vector. Finally, a classifier is trained on top of the standardized vector. Related to the late fusion (VOTING): each modality such as the image and text are propagated through their classifier. Finally, the output probabilities of both models are averaged, and the class is selected using the maximum value. It is worth mentioning that a third type is known as Hybrid Fusion (GMU) exists and can be used to support multiple languages(input text belongs to multiple languages). This type is usually trained using the best features per modalities.

CHAPTER V
EXPERIMENTS

In this section, we evaluate the proposed deep learning approach and some machine learning algorithms for processing multimodal natural disaster classification tasks. It is worth mentioning that building and training models from scratch are a very challenging stage, as it requires a large number of training samples, and strong processors and GPUs. Instead, if we retrain ready-trained models using the proposed dataset, such process will be much easier, and a normally equipped computer will also be suitable for training such model.

In this work, achievement results were examined by retraining formerly trained models in our dataset. Our experiments were conducted on a machine equipped with the Nvidia GTX1650 GPU with Intel(R) Core (TM) i7-10750H CPU and 16 GB of RAM running the Windows 10 Enterprise operating system. All code implemented and libraries are used are implemented in Python. Google Colab environment is used for implementing the code. In this study, which we developed using Python programming language, the Keras package allows us to perform some preprocessing steps. Keras allows us to use datasets by a specific index structure.

In the study, performance measurements were carried out with three separate classifications; only tweet text, only tweet image, and tweet text and image together. The data preprocessing steps and training performance results are described below. The performance of the trained models is evaluated by multiple evaluation matrices such as accuracy, precision, recall, and F1 Score.

5.1. Dataset

To make a multi-modal classification for natural disasters, in this study, data consisting of Turkish language texts and their related images were collected from Twitter. Unfortunately, until the time of this study, there is no Turkish tweet text and image dataset. Hence, we are introducing the first dataset. The data to be used in the analysis of whether it is relevant to natural disasters is taken from Twitter, which is a very widely used microblog. Figure 5.2 shows the hashtags that are used to identify a disaster event. To create a new dataset;

deprem(earthquake), yangın(fire), trafik kazası(traffic accident), müsilaj(sea saliva), and sağanak(downpour) keywords were used. Five different separate datasets with text and image data are created related to the mentioned natural disasters. However, as data may contain some meaningless words and symbols, the cleanup is performed afterward. Then, three annotators first read each tweet and view the related image(s), and then independently judged whether the texts and images are related to the natural disaster or not. The majority judgment result was applied for the final labeling. Manual annotation instructions are shown in Table 5.1 and Table 5.2 shows samples of collected tweets with keywords and classes.

Note that some people are sharing some irrelevant tweets during disasters with keywords and hashtags related to the disaster to increase the number of views and shares. Hence, it is important to distinguish tweets about the disaster. A sample of the collected relevant and irrelevant tweets (text and image) for different natural disasters are shown in Figure 5.1. One of the main objectives of this thesis is to improve the multimodal classification of informative tweets during natural disasters with combinations of text and image characteristics. In the natural disaster classifications made so far, the focus has been on text characteristics in general and no Turkish sources made with image characteristics have been found in the literature.

Table 5.1. Manual annotation tasks and instructions

Task	Instruction
Relevant	If the text and image of the input tweet are related to one of the selected natural disasters (#deprem, #yangın, #sağanak, #trafik accident, #müsilaj), the tweet will be tagged about that natural disaster.
Irrelevant	If the text and image of the input tweet are not related to one of the selected natural disasters (#deprem, #yangın, #sağanak, #trafik accident, #müsilaj), the tweet will be considered and tagged as irrelevant.

Table 5.2. Sample of tweets with keywords and classes

Keyword	Tweet	Meaning	Class
Deprem	Bi an uçağın gürültüsü zannettim ağaçlar sallandı #deprem	At one moment I thought it was the noise of the plane, the trees shook #deprem	Irrelevant
Deprem	#SONDAKİKA #deprem Son dakika... Ege Denizi'nde korkutan deprem	Last-minute... Terrifying earthquake in the Aegean Sea	Relevant
Yangın	Her tünelin sonundaki ışığın aslında yangın olmasından sıkıldım	I'm tired of the light at the end of every tunnel being a fire.	Irrelevant
Yangın	@UrfahaberNet : Şanlıurfa Yarı Açık Cezaevi'nde yangın: İtfaiye müdahale etti	@UrfahaberNet : Fire at Sanliurfa Semi-Open Prison: Fire brigade intervenes	Relevant
Trafik Kazası	Ceyhan Zabıta Müdürü Ali Akar Geçirmiş Olduğu Trafik kazası Nedeniyle Şehir Hastanesi Yoğun Bakımda Arkadaşıma Acil Şifalar diliyorum	Ceyhan Police Chief Ali Akar I wish urgent healing to my friend in the Intensive Care Unit of The City Hospital due to the traffic accident he has had	Irrelevant
Trafik Kazası	Konya'da trafik kazası: 1 yaralı	Konya traffic accident: 1 injured	Relevant
Müsilaj	Yalan Uydurmanın ve Yaymanın Tedavisi Yok. Aşısı Yok. Yalan Siyaseti Hakkında "Siyasi Müsilaj" İlan Ettik.	There's No Cure for Lying and Spreading. There's no vaccine. We declared "Political Sea saliva" about the Politics of Lies.	Irrelevant
Müsilaj	İzmit Körfezi'nin bazı kesimlerinde müsilaj yoğunluğu azaldı	Sea saliva density decreased in parts of Izmit Bay	Relevant
Sağanak	ORDU'DA SAĞANAK YAĞIŞ; YOLLAR GÖLE DÖNDÜ!	DOWNPOUR IN ORDU; THE ROADS ARE TURNED INTO LAKES!	Relevamt
Sağanak	Dubai drone yağmurlari ile sağanak yağmur ile çölü ormana çevirmeye çalışıyor. Yaşanacak ülke çevre oluşturmak bina köprü dikmek çok kolaydır kısa sürelidir ama insanın zihnine insanlığı, düşünceyi, doğruyu, bilgiyi inşaa etmek kadar zor birşey yoktur.	Dubai is trying to turn the desert into a jungle with drone rains. It is very easy to build a building bridge to create a country of life, but there is nothing more difficult than building humanity, thought, truth, knowledge into one's mind.	Irrelevant

Figure 5.1. Sample images along with their text from collected tweets

It is known that one of the factors to develop a deep learning model, is to have non-overlapping train/dev/test sets, where the used dataset is often randomly split into train/dev/test sets. As shown in Table 5.3, in our study, we made the division process at a rate of 70%, 15%, and 15%, respectively. Note that these are subsets of the datasets reported in Table 5.4.

Table 5.3 Number of samples in different splits of our datasets.

	Train (70%)		Dev (15%)		Test (15%)		Total (100%)	
	Text	**Image**	**Text**	**Image**	**Text**	**Image**	**Text**	**Image**
Relevant	6015	5007	1288	755	1288	755	8591	6517
Irrelevant	5894	4512	263	967	1263	967	8420	6447

Figure 5.2. Sample of hashtags used for data collection.

Table 5.4 Number of samples for each sub-dataset.

Disaster	Datasets	# Relevant Tweets	#Irrelevant Tweets
Deprem	Dataset 1	2596	2400
Yangın	Dataset 2	1450	1448
Trafik Kazası	Dataset 3	1718	1692
Müsilaj	Dataset 4	800	780
Sağanak	Dataset 5	2029	2100

In the following experiments, the performance of ensemble systems and state-of-the-art CNN models that were originally developed for unimodal and multimodal classification was examined using all the five datasets that have been created. To provide the quality of the achieved results, five datasets, which contain samples of balanced relevant and irrelevant texts and images sequentially, were used to compare the performance of the mentioned models.

5.2. Unimodal Classification

In this section, Unimodal image and text approaches were trained separately for each natural disaster (deprem-earthquake, yangın- fire, trafik kazası-traffic accident, müsilaj-mucilaj, sağanak- downpour). Table 5.3 shows samples of tweets with keywords and classes and the number of samples are given in detail in Table 5.4.

Experiment 1. Performance of Unimodal Classification (Image and Text)

For the Unimodal image experiment, we used five different CNN architectures. Figure 5.3 and Table 5.5 reports the accuracy and the F1 Score for the tested CNN models. The main aim of this experiment is to choose the best architecture that will be later adapted and used for multimodal fusion. As a result, although all models achieved almost similar performance, "Mouzannar's CNN model1" performed marginally better than other models in the validation set, and were also the fastest model for training and forecasting.

Table 5.5. Performance of the selected CNN models when used for image classification (unimodal).

State of Art CNN Models / Dataset	Dataset1		Dataset2		Dataset3		Dataset4		Dataset5	
	Acc	F1 Score	Acc	F1 Score	Acc	F1 Score	Acc	F1 Score	Acc	F1 Score
CNN_Model1 (Mouzannar,2018)	**0.8548**	**0.852**	**0.8522**	**0.8501**	**0.8578**	**0.8552**	**0.8601**	**0.8579**	**0.8512**	**0.8479**
CNN_Model2 (Alam,2020)	0.8321	0.8309	0.8295	0.8287	0.8342	0.8321	0.8344	0.8323	0.8301	0.8279
VGG19	0.7183	0.7169	0.717	0.7157	0.7219	0.7203	0.7238	0.7217	0.7141	0.7119
ResNet50	0.7333	0.7315	0.7319	0.7302	0.7385	0.7364	0.7407	0.7386	0.7311	0.7381
Inception_ResnetV2	0.8313	0.8316	0.8291	0.8275	0.8342	0.8327	0.8369	0.8342	0.8301	0.8279

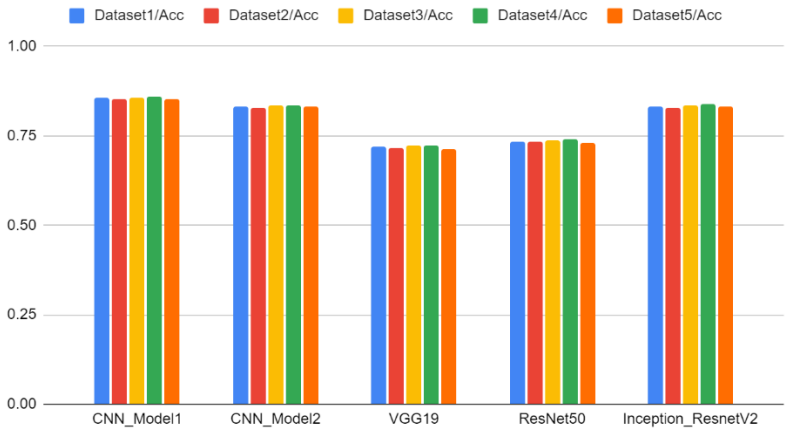

Figure 5.3. Performance of the selected CNN models when used for image classification (unimodal).

We applied LSTM and Kim's CNN architecture for the text model (Kim, 2014). We compared pre-trained Glove, Word2Vec, and BERT with three different word embeddings. Table 5.6 summarizes the performance results of CNN and an LSTM model, which is reserved for training and tested in a validation set. The number of filters did not have a significant effect on accuracy; the default parameters of the network were effective for this text model. Word2Vec outperformed GloVe by more than 1% in the validation set, but BERT marginally outperformed both Word2Vec and GloVe.

Table 5.6. Performance of the classifiers using multiple embedding approach.

Classifier	Word Embedding	Accuracy	Precision	Recall	F1-Score
Proposed Model_LSTM	Glove	0.8644	0.8581	0.8704	0.8642
	Word2Vec	0.8721	0.8643	0.8769	0.8701
	BERT	**0.8795**	**0.8702**	**0.8845**	**0.8772**
CNN(Kim,2014)	Glove	0.8498	0.8421	0.8546	0.8483
	Word2Vec	0.8547	0.8495	0.8592	0.8543
	BERT	**0.8576**	**0.8531**	**0.8607**	**0.8568**

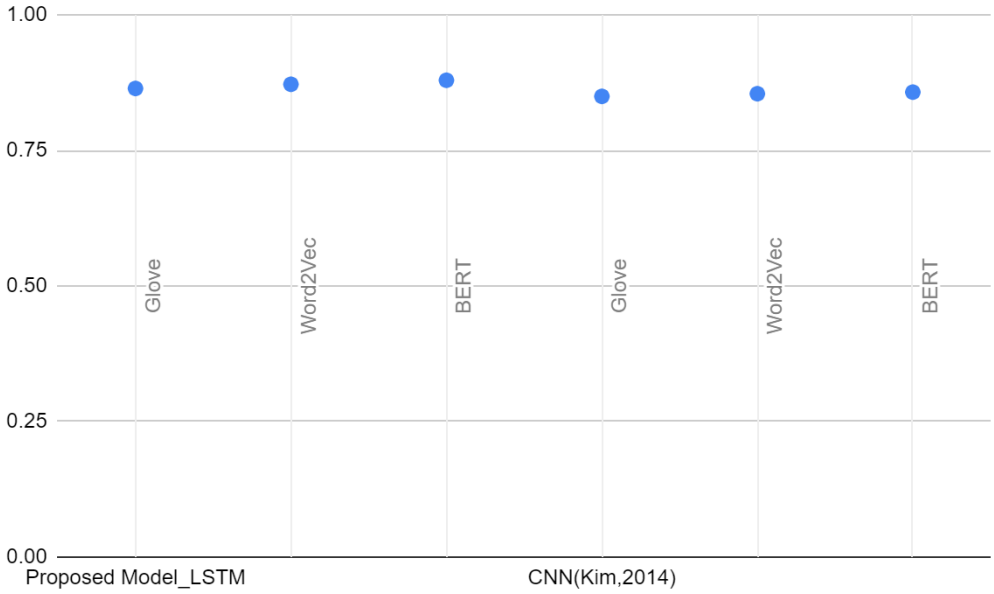

Figure 5.4. Performance of the classifiers using multiple embedding approach.

Experiment 2. The Effects of Preprocessing the Tweets on the Performance of the ensemble and deep learning models for unimodal text classification.

The performance of the possible ensemble systems that are applied using 14 different machine learning algorithms on Turkish tweets. The results are shown in Table 5.8 and Figure 5.5. Based on the results, Gradient Boosting has achieved the best performance.

Table 5.7 shows that preprocessing affects the performance of the classifiers. In addition to this, the objective of this experiment is to find how the performance can be affected by either preprocessing the data or using it directly.

Table 5.7. Accuracy of classifiers with/without preprocessing**.**

Classifier	**Accuracy/ With Preprocessing**	**Accuracy/ Without Preprocessing**	**Improvement %**
Proposed Model_LSTM	0.8721	0.8542	2,08
CNN(Kim,2014)	0.8547	0.8435	1,3

Table 5.8 Accuracy of ensemble systems with/without the preprocessing

Classifiers	Accuracy/ Without Preprocessing	Accuracy / With Preprocessing	Improvement %
Nearest_Neighbors	0.6576	**0.6593**	0,26
Linear_SVM	0.5762	**0.5881**	2,02
Polynomial_SVM	0.5661	**0.5814**	2,63
RBF_SVM	0.5203	**0.6000**	13,28
Gaussian_Process	0.5457	**0.6322**	13,68
Gradient_Boosting	0.7980	**0.8051**	0,88
Decision_Tree	0.6610	**0.7016**	5,79
Extra_Trees	0.7796	**0.7898**	1,29
Random_Forest	0.7339	**0.7593**	3,35
Neural_Net	0.6136	**0.6779**	9,49
AdaBoost	0.7135	**0.7356**	3,00
Naive_Bayes	0.5474	**0.5831**	6,12
QDA	0.5711	**0.6051**	5,62
SGD	0.5406	**0.5458**	0,95

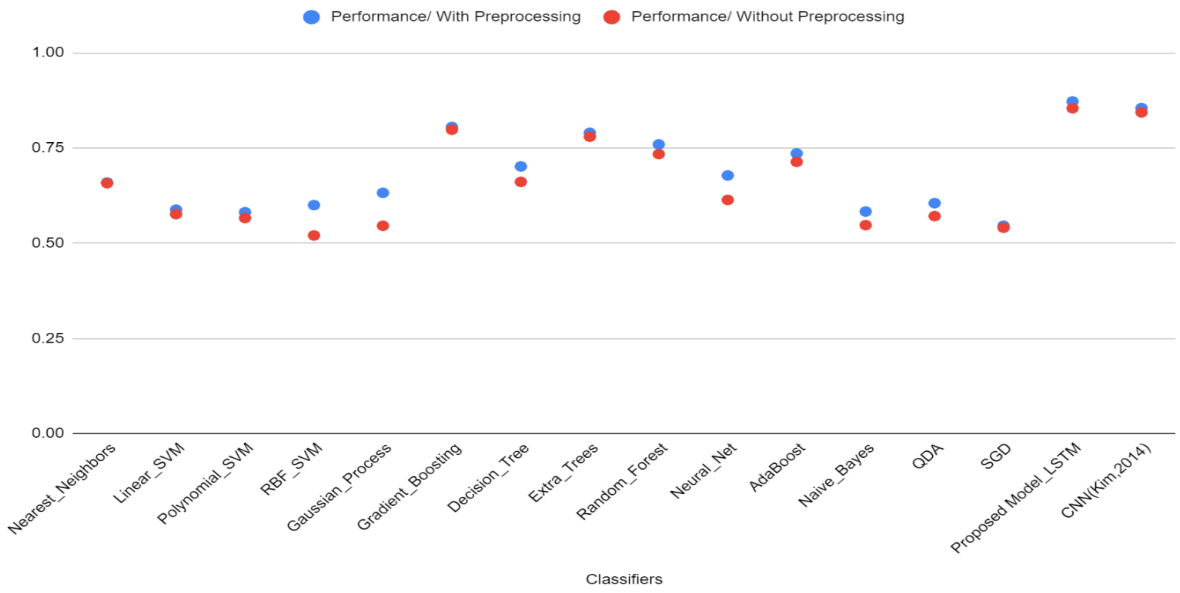

Figure 5.5. The accuracy of all models with and without the preprocessing operation.

5.3. Multimodal Classification

Experiment 3. Performance of the State of Art Deep Learning Models

Based on the results using the validation set, we selected Mouzannar's CNN model and the BERT- LSTM text model for the multimodal classifiers. First of all, we evaluated the late fusion approach. For rule-based classifiers, text data was trained with LSTM and image data with CNN classifiers to implement the weighted maximum decision rule model. Performance results are shown in Table 5.9 and Figure 5.6.

Table 5.9. Accuracy, precision, recall, and F1 score for the proposed model

BERT LSTM- CNN	Dataset1	Dataset2	Dataset3	Dataset4	Dataset5
Accuracy	91.87	88.46	90.87	91.07	91.23
Precision Score	90.34	81.81	89.12	89.63	90.03
F1 Score	91.28	87.8	88.89	90.44	91.05
Recall Score	92.25	90.73	91.16	91.28	92.11

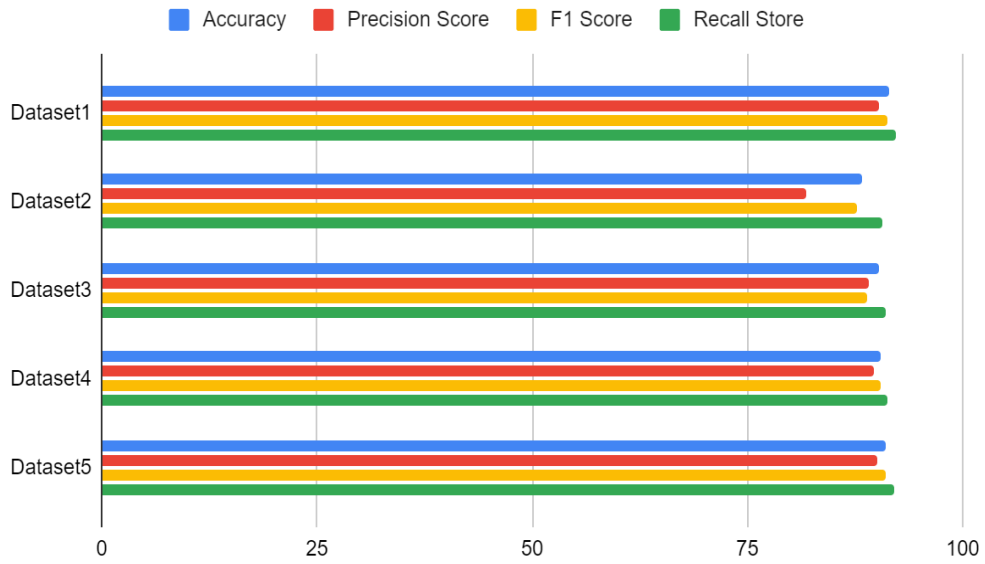

Figure 5.6. Accuracy, precision, recall, and F1 score for the proposed model.

Table 5.10 shows the performance comparisons of three states of art deep learning methods such as CNN(Word2Vec) – CNN (Nyugen,2017), LSTM (Word2Vec)-CNN(Mouzannar,2018), and Proposed Model LSTM(BERT) – CNN. Based on the results, the proposed method has the best performance.

Table 5.10. Performance comparisons of the state-of-art deep learning models

Models	Datasets	Precision	Recall	F1 Score	Accuracy
CNN (Word2Vec) – CNN (Nyugen, 2017)	Dataset1	85.93	87.13	86.52	86.47
	Dataset2	84.42	85.82	85.11	85.12
	Dataset3	85.08	86.14	85.60	85.62
	Dataset4	84.63	85.94	85.27	85.44
	Dataset5	85.15	86.36	85.75	85.93
LSTM (Word2Vec)- CNN (Mouzannar, 2018)	Dataset1	90.72	91.95	91.33	91.46
	Dataset2	87.45	88.93	88.18	88.15
	Dataset3	89.77	91.02	90.39	90.38
	Dataset4	89.69	90.94	90.31	90.47
	Dataset5	90.29	91.71	90.99	91.07
LSTM (BERT)- CNN Proposed Model	Dataset1	90.34	92.25	91.28	91.87
	Dataset2	81.81	90.73	87.80	88.46
	Dataset3	89.12	91.16	88.89	90.87
	Dataset4	89.63	91.28	90.44	91.07
	Dataset5	90.03	92.11	91.05	91.23

In this thesis, we discover that accuracy from classification models for the late fusion of text and image performed better and made the classification more effective. In Table 5.11 we proffer the performance results achieved for different modalities. Text-only unimodal performs better than image-only unimodal. Particularly, multimodal models provide a further performance improvement.

Table 5.11. Performance comparison of different modalities

Training Modal	Modality	Accuracy	Precision	Recall	F1 Score
Unimodal	Image	85.48	84.87	86.55	85.20
	Text	87.95	87.02	88.45	87.92
Multimodal	Text + Image	91.87	90.34	92.25	91.28

CHAPTER IX
CONCLUSION AND FUTURE WORKS

Information collected from different data modalities obtained from social media for disaster response is very useful. Although the images shared on social media contain significantly beneficial information, the studies carried out so far have focused on text analysis. In this thesis, both text and images taken from social media are classified, and a new model is created. In particular, deep learning architectures were used to make the multimode classification system. In the experimental section that was carried out on the proposed datasets created on real disasters, the usefulness of the method proposed in the thesis is revealed. In this sense, we also state that improvements can be made for future studies in our system, where texts and images are classified multimodal in Turkish.

The late fusion was used to achieve multimodal classification; Also, a pre-trained BERT - LSTM model is used in text modal while a pre-trained CNN model is used in visual modal. Concatenating both inputs in a multimodal learning architecture achieved an accuracy of 91.87%. It can be seen from the accuracy result that a multimodal deep learning method can improve the classification accuracy of disaster events, and classified data may help authorities to make the most suitable decisions on time.

REFERENCES

Alam F, Ofli F, Imran M (2020) Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of hurricanes harvey, irma, and maria. Behaviour & Information Technology 39(3):288–318.

Alam, F.; Sajjad, H.; Imran, M.; and Ofli, F. 2021. CrisisBench:Benchmarking Crisis-related Social Media Datasets for Humanitarian Information Processing. InICWSM

Alharbi, A.; and Lee, M. 2019. Crisis Detection from ArabicTweets. InWorkshop on Arabic Corpus Ling., 72–79

Basu M, Shandilya A, Khosla P, Ghosh K, Ghosh S (2019) Extracting resource needs and availabilities from microblogs for aiding post-disaster relief operations. IEEE Transactions on Computational Social Systems 6(3):604–618

Breiman L (2001) Random forests. Machine learning 45(1):5–32

Bouslimi, R., Ayadi, M.G. & Akaichi, J., (2017). Semantic Medical Image Retrieval in A Medical Social Network. Social Network Analysis and Mining, 7(2), 1-11. doi: https://doi.org/10.1007/s13278-016-0420-3.

Caragea C, Silvescu A, Tapia AH (2016) Identifying informative messages in disaster events using convolutional neural networks. In: International Conference on Information Systems for Crisis Response and Management, pp 137–147.

Caragea C, Squicciarini AC, Stehle S, Neppalli K, Tapia AH (2014) Mapping moods: Geo-mapped sentiment analysis during hurricane sandy. In: ISCRAM

Castillo C (2016) Big crisis data: social media in disasters and time-critical situations. Cambridge University Press

Cees G. M. Snoek, Marcel Worring, and Arnold W. M. Smeulders, “Early versus Late Fusion in Semantic Video Analysis”, in Proceedings of the ACM International Conference on Multimedia, Singapore, 2005, pp. 399-402

Crowe, Adam. “Emergency Management Websites.” Crisis Response Journal 4, no. 4 (accessed December 30, 2010).

Demirtaş, Z. G. & Demirtaş, İ. (2017). KRİZ DÖNEMLERİNDE SOSYAL MEDYA KULLANIMI:15 TEMMUZ DARBE (KALKIŞMA) GİRİŞİMİ SONRASINDA TÜRKİYE’DEKİ BAKANLAR KURULU ÜYELERİNİN TWİTTER KULLANIMI ÜZERİNE BİR İNCELEME . Süleyman Demirel Üniversitesi Vizyoner Dergisi , 8 (19) , 137-146 . DOI: 10.21076/vizyoner.343028

Dunning T (1993) Accurate methods for the statistics of surprise and coincidence. Computational Linguistics 19(1):61–74. https://www.aclweb.org/anthology/J93-1003.

E. Alpaydın (2018). “Classifiying Multimodal Data,” In S, Oviatt, B. Schuller, P. Cohen, D. Sonntag, G. Potamianos, and A. Krueger, editors, The Handbook of Multimodal-Multisensor Interfaces, Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition. Chapter 2, pp. 49-69, Morgan & Claypool Publishers, San Rafael, CA

E. Cambria, B. White, 2014, Jumping NLP curves: A review of natural language processing research, IEEE Comput. Intell. Mag. 9 48–57

Enenkel M, Saenz SM, Dookie DS, Braman L, Obradovich N, Kryvasheyeu Y (2018) Social media data analysis and feedback for advanced disaster risk management. CoRR abs/1802.02631 arXiv:1802.02631

Fersini, E., Messina, E. and Pozzi, F.A., 2017. Earthquake management: a decision support system based on natural language processing. Journal of Ambient Intelligence and Humanized Computing, 8(1), pp.37-45.

Freund Y, Schapire RE et al (1996) Experiments with a new boosting algorithm. In: Icml, Bari, Italy, vol. 96, pp 148–156.

Hadi S Jomaa, Yara Rizk, and Mariette Awad. 2016. Semantic and Visual Cues for Humanitarian Computing of Natural Disaster Damage Images. In 12th Int. Conf. on Signal-Image Technology & Internet-Based Systems. 404–411.

Hassan, S.Z.; Ahmad, K.; Hicks, S.; Halvorsen, P.; Al-Fuqaha, A.; Conci, N.; Riegler, M. Visual Sentiment Analysis from Disaster Images in Social Media. arXiv 2020, arXiv:2009.03051.

Houston JB, Hawthorne J, Perreault MF, Park EH, Goldstein Hode M, Halliwell MR, Turner McGowen SE, Davis R, Vaid S, McElderry JA, et al. (2015) Social media and disasters: a functional framework for social media use in disaster planning, response, and research. Disasters 39(1):1–22.

DataReportal (2021), “Digital 2021 Global Digital Overview,” retrieved from https://datareportal.com/reports/digital-2021-global-digital-overview.

DOMO (2020), “Data Never Sleeps 8.0”, https://www.domo.com/assets/downloads/18_domo_data-never-sleeps-6+verticals.pdf

Imran M, Castillo C, Diaz F, Vieweg S (2015) Processing social media messages in mass emergency: A survey. ACM Computing Surveys (CSUR) 47(4):67

Jain, P.; Ross, R.; and Schoen-Phelan, B. 2019. Estimating Distributed Representation Performance in Disaster-Related SocialMedia Classification. InASONAM

Java, Akshay, "Why we Twitter: Understanding microblogging usage and communities", Proceeding of the 9th Web KDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. ACM,2007

Kiatpanont, R., Tanlamai, U. and Chongstitvatana, P., 2016. Extraction of actionable information from crowdsourced disaster data. Journal of emergency management, 14(6), pp.377-390.

Kryvasheyeu Y, Chen H, Obradovich N, Moro E, Van Hentenryck P, Fowler J, Cebrian M (2016) Rapid assessment of disaster damage using social media activity. Science advances 2(3):e1500779

Loodos., "loodos/bert-base-turkish-uncased hugging face," https://github.com/Loodos/turkish-languagemodels, Aug. 2020.

Luhn HP (1957) A statistical approach to mechanized encoding and searching of literary information. IBM Journal of research and development 1(4):309–317.

Madichetty S et al (2018) Re-ranking feature selection algorithm for detecting the availability and requirement of resources tweets during disaster. International Journal of Computational Intelligence & IoT, 1(2).

Madichetty S, Sridevi M (2019) Detecting informative tweets during disaster using deep neural networks. In: 2019 11th International Conference on Communication Systems & Networks (COMSNETS), pp 709–713. IEEE.

Madichetty S, Sridevi M (2019) Disaster damage assessment from the tweets using the combination of statistical features and informative words. Social Network Analysis and Mining 9(1):42.

Mitchell JT, Thomas DeborahSK, Hill AA, Cutter SL (2000) Catastrophe in reel life versus real life: Perpetuating disaster myth through hollywood films. International Journal of Mass Emergencies and Disasters 18(3):383–402.

Mouzannar, H., Rizk, Y. and Awad, M. (2018). Damage Identification in Social Media Posts using Multimodal Deep Learning. Proceedings of the 15th International Conference on Information Systems for Crisis Response and Management (ISCRAM), Rochester, 529-543.

N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, 2014, Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 1929–1958.

Nguyen DT, Ofli F, Imran M, Mitra P (2017) Damage assessment from social media imagery data during disasters. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, pp 569–576. ACM.

Pandey, N. and Natarajan, S., 2016, September. How social media can contribute during disaster events? Case study of Chennai floods 2015. In Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on (pp. 1352-1356). IEEE

Purohit H, Castillo C, Diaz F, Sheth A, Meier P (2014) Emergency-relief coordination on social media: Automatically matching resource requests and offers. First Monday, 19(1)

Rudra K, Ganguly N, Goyal P, Ghosh S (2018) Extracting and summarizing situational information from the twitter social media during disasters. ACM Transactions on the Web (TWEB) 12(3):17

Rudra K, Ganguly N, Goyal P, Ghosh S (2018) Extracting and summarizing situational information from the twitter social media during disasters. ACM Transactions on the Web (TWEB) 12(3):17

Rudra K, Ghosh S, Ganguly N, Goyal P, Ghosh S (2015) Extracting situational information from microblogs during disaster events: a classification-summarization approach. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp 583–592. ACM.

Rudra K, Ghosh S, Ganguly N, Goyal P, Ghosh S (2015) Extracting situational information from microblogs during disaster events: a classification-summarization approach. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp 583–592. ACM.

Rudra K, Sharma A, Ganguly N, Imran M (2017) Classifying information from microblogs during epidemics. In: Proceedings of the 2017 International Conference on Digital Health, pp 104–108. ACM

S. Hochreiter, J. Schmidhuber, 1997, Long short-term memory Neural Comput. 9 1735–1780.

Sakai, T., Tamura, K., Kitakami, H. and Takezawa, T., 2017, November. Photo image classification using pre-trained deep network for density-based spatiotemporal analysis system. In Computational Intelligence and Applications (IWCIA), 2017 IEEE 10th International Workshop on (pp. 207-212). IEEE

Sarter NB, Woods DD (1991) Situation awareness: A critical but ill-defined phenomenon. Int J Aviat Psychol 1(1):45–57

Tanev, H., Zavarella, V. and Steinberger, J., 2017. Monitoring disaster impact: detecting micro-events and eyewitness reports in mainstream and social media. In Proceedings of the 14th ISCRAM Conference – Albi, France.

The Rise of Social Media. https://ourworldindata.org/rise-of-social-media

Tran N, Schneider J G, Weber I, Qin A K. "Hyperparameter optimization in classification: To-do or not-to-do". Pattern Recognition, 103, 107245, 2020.

Vieweg S, Castillo C, Imran M (2014) Integrating social media communications into the rapid assessment of sudden onset disasters. In: International Conference on Social Informatics, vol 8851. Springer, pp 444–461.

Wiegmann, M.; Kersten, J.; Klan, F.; Potthast, M.; and Stein, B.2020. Analysis of Detection Models for Disaster-Related Tweets.InISCRAM.

Yoshua Bengio. Deep Learning of Representations for Unsupervised and Transfer Learning. In JMLR: Workshop and Conference Proceedings, volume 7, pages 1–20, jun 2011. 28, 37

Young S R, Rose D C, Karnowski T P, Lim S H, Patton R M. "Optimizing deep learning hyperparameters through an evolutionary algorithm". Workshop on Machine Learning in HighPerformance Computing Environments, November 2015.