Rumor Classification through a Multimodal Fusion Framework and Ensemble Learning

Abderrazek Azri, Cécile Favre, Nouria Harbi, Jérôme Darmont, Camille Noûs

Rumor Classification through a Multimodal Fusion Framework and Ensemble Learning

Abderrazek Azri[1*], Cécile Favre[1], Nouria Harbi[1], Jérôme Darmont[1] and Camille Noûs[2]

[1]Université de Lyon, Lyon 2, UR ERIC,
5 avenue Pierre Mendès France, 69676 Bron Cedex, France.
[2]Université de Lyon, Lyon 2, Laboratoire Cogitamus, France.

*Corresponding author(s). E-mail(s): a.azri@univ-lyon2.fr;
Contributing authors: cecile.favre@univ-lyon2.fr;
nouria.harbi@univ-lyon2.fr; jerome.darmont@univ-lyon2.fr;
camille.nous@cogitamus.fr;

Abstract

The proliferation of rumors on social media has become a major concern due to its ability to create a devastating impact. Manually assessing the veracity of social media messages is a very time-consuming task that can be much helped by machine learning. Most message veracity verification methods only exploit textual contents and metadata. Very few take both textual and visual contents, and more particularly images, into account. Moreover, prior works have used many classical machine learning models to detect rumors. However, although recent studies have proven the effectiveness of ensemble machine learning approaches, such models have seldom been applied. Thus, in this paper, we propose a set of advanced image features that are inspired from the field of image quality assessment, and introduce the Multimodal fusiON framework to assess message veracIty in social neTwORks (MONITOR), which exploits all message features by exploring various machine learning models. Moreover, we demonstrate the effectiveness of ensemble learning algorithms for rumor detection by using five metalearning models. Eventually, we conduct extensive experiments on two real-world datasets. Results show that MONITOR outperforms state-of-the-art machine learning baselines and that ensemble models significantly increase MONITOR's performance.

Keywords: Social networks, Rumor verification, Image features, Machine learning, Ensemble learning

1 Introduction

After more than two decades of existence, social media platforms have attracted a large number of users. They enable the diffusion of information in real-time, albeit regardless of its credibility, for two main reasons. First, there is a lack of a means to verify the veracity of contents transiting on social media. Second, users often publish messages without verifying information validity and reliability. Consequently, social networks, and particularly microblogging platforms, are a fertile ground for spreading rumors.

Widespread rumors can pose a threat to the credibility of social media and cause harmful consequences in real life. Thus, the automatic assessment of information credibility on microblogs that we focus on is crucial to provide decision support to, e.g., fact checkers. This task requires to verify the truthfulness of messages related to a particular event and return a binary decision stating whether the message is authentic.

In the literature, most automatic rumor detection approaches address the task as a classification problem. They generally extract features from two aspects of messages: textual content (Pérez-Rosas, Kleinberg, Lefevre, & Mihalcea, 2018) and social context (L. Wu & Liu, 2018). However, the multimedia content of messages, particularly images that present a significant set of features, are little exploited.

In this paper, we second the hypothesis that the use of image properties is important in rumor verification. Images indeed play a crucial role in the news diffusion process. For example, in the dataset collected by Jin, Cao, Zhang, Zhou, and Tian (2017), the average number of messages with an attached image is more than eleven times that of plain text messages.

Figure 1 shows two sample rumors posted on Twitter. In Figure 1a, it is hard to assess veracity from the text, but the likely-manipulated image hints at a rumor. In Figure 1b, it is hard to assess veracity from both the text or the image because the image has been taken out of its original context.

(a) Black clouds in New York City before Sandy!!!

(b) NepalEarthquake 4Years old boy protect his little sister. make me feel so sad

Fig. 1 Two sample rumors posted on Twitter

Furthermore, most of the literature focuses on features to train a wide range of machine learning (Volkova & Jang, 2018) and deep learning (Wang et al., 2018) methods. However, although recent studies demonstrate the effectiveness of ensemble learning (Gutierrez-Espinoza, Abri, Namin, Jones, & Sears, 2020), such models are not applied for rumor detection.

Based on the above observations, we aim to leverage all the modalities of microblog messages for verifying rumors, that is, features extracted from the textual and social context content of messages, and up to now unused visual and statistical features derived from images. Consequently, all types of features must be fused to allow a supervised machine learning classifier to evaluate the credibility of messages. Moreover, motivated by the recent research on ensemble learning to classification problems (Pang, Xue, & Namin, 2016), we design various metalearning models to investigate the performance of ensemble learning for rumor classification.

Our contribution is threefold. First, we propose the use of a set of image features inspired from the field of Image Quality Assessment (IQA) and we show that they contribute very effectively to the verification of message veracity. These metrics estimate the rate of noise and quantify the amount of visual degradation of any type in an image. They are proven to be good indicators for detecting fake images, even those generated by advanced techniques such as Generative Adversarial Networks (GANs) (Goodfellow et al., 2014). To the best of our knowledge, we are the first to systematically exploit this type of image features to check the veracity of microblog posts.

Second, we detail the Multimodal fusiON framework to assess message veracIty in social neTwORks (MONITOR) (Azri, Favre, Harbi, Darmont, & Noûs, 2021b), which exploits all types of message features and leverages four machine learning models that provide explainability and interpretability about the taken decisions.

Third, we demonstrate the benefit of ensemble learning, by developing five metalearning models (soft and weighted average voting, stacking, blending, and super learner ensemble) that exploit the above four machine learning models, and we compare their performance with MONITOR's. To the best of our knowledge, we are the first to apply metalearning models for tackling the rumor detection task.

Eventually, we conduct extensive experiments two real-world datasets to show the effectiveness of our rumor detection approach. MONITOR indeed outperforms all state-of-the-art machine learning baselines with an accuracy and F1-score of up to 96% and 89% on the MediaEval benchmark (Boididou et al., 2015) and the FakeNewsNet dataset (Shu, Mahudeswaran, Wang, Lee, & Liu, 2018), respectively. Furthermore, all metalearning algorithms notably increase MONITOR's performance.

The remainder of this paper is organized as follows. In Section 2, we review all the research related to our problem. In Section 3, we detail MONITOR and especially its feature extraction and selection. In Section 4, we present and

comment on the experimental results that we achieve with respect to state-of-the-art methods. In Section 5, we investigate and discuss the performance of ensemble models. Finally, in Section 6, we conclude this paper and outline future research.

2 Related Works

Related work can be divided into the following categories:

1. non-image features and image features that are essential for checking the veracity of microblog posts,
2. background information regarding ensemble learning models and their usage for rumor classification.

2.1 Non-image Features

Studies in the literature present a wide range of non-image features. These features may be divided into two subcategories, textual features and social context features. To classify a message as fake or real, Castillo, Mendoza, and Poblete (2011) capture prominent statistics in tweets, such as count of words, capitalized characters and punctuation. Beyond these features, lexical words expressing specific semantics or sentiments are also counted. Many sentimental lexical features are proposed (Kwon, Cha, Jung, Chen, & Wang, 2013), which utilize a sentiment tool called the Linguistic Inquiry and Word Count (LIWC) to count words in meaningful categories.

Other works exploit syntactic features, such as the number of keywords, the sentiment score or polarity of the sentence. Features based on topic models are used to understand messages and their underlying relations within a corpus. K. Wu, Yang, and Zhu (2015) train a Latent Dirichlet Allocation model (Blei, Ng, & Jordan, 2003) with a defined set of topic features to summarize semantics for detecting rumors.

The social context describes the propagating process of a rumor (Shu, Wang, & Liu, 2018). Social network features are extracted by constructing specific networks, such as diffusion (Kwon et al., 2013) or co-occurrence networks (Ruchansky, Seo, & Liu, 2017).

Recent approaches detect fake news based on temporal-structure features. Kwon, Cha, and Jung (2017) studied the stability of features over time and found that, for rumor detection, linguistic and user features are suitable for early-stage, while structural and temporal features tend to have good performance in the long-term stage.

2.2 Image Features

Although images are widely shared on social networks, their potential for verifying the veracity of messages in microblogs is not sufficiently explored. Morris, Counts, Roseway, Hoff, and Schwarz (2012) assume that the user's profile image has an important impact on information credibility. Images attached in

messages bear very basic features. K. Wu et al. (2015) define a feature called "has multimedia" to mark whether the tweet has any picture, video or audio attached. A. Gupta, Lamba, Kumaraguru, and Joshi (2013) propose a classification model to identify fake images on Twitter during Hurricane Sandy. However, their work is still based on textual content features.

To automatically predict whether a tweet that shares multimedia content is fake or real, Boididou et al. (2015) propose the Verifying Multimedia Use (VMU) task. Textual and image forensics (Li, Li, Yang, & Sun, 2014) features are used as baseline features for this task. They conclude that Twitter media content is not amenable to image forensics and that forensics features do not lead to consistent VMU improvement (Boididou et al., 2018).

2.3 Ensemble Learning

Ensemble learning refers to the generation and combination of multiple inducers to solve a particular machine learning task. The intuitive explanation for the ensemble methodology stems from human nature. Often, decision making by a group of individuals results in more accurate, useful or correct outcome than a decision made by any one member of the group. This is generally referred to as the wisdom of the crowd (Surowiecki, 2005). Using ensemble learning, the performance of poorly performing classifiers can be improved by creating, training and combining the output of multiple classifiers and thus result in a more robust classification. There are three main approaches for developing an ensemble learner (Zhang & Ma, 2012):

- *boosting* uses homogeneous-base models trained sequentially;
- *bagging* (Bootstrap AGGregatING) uses homogeneous-base models trained in parallel;
- *stacking* uses mostly heterogeneous-base models trained in parallel and combined using a metamodel.

By averaging (or voting) the output produced by the pool of classifiers, ensemble methods provide better predictions and avoid overfitting. Another reason that contributes to the better performance of ensemble learning is its ability in escaping from local minimums. By using multiple models, the search space becomes wider and the chance for finding a better output becomes higher (Sagi & Rokach, 2018).

Recently ensemble learning methods have shown good performance in various applications, including solar irradiance prediction (J. Lee, Wang, Harrou, & Sun, 2020), slope stability analysis (Pham, Kim, Park, & Choi, 2021), natural language processing (Sangamnerkar, Srinivasan, Christhuraj, & Sukumaran, 2020), malware detection (D. Gupta & Rani, 2020), COVID-19 detection (Singh, Kaur, Singh, & Dhiman, 2021), movie success detection (K. Lee, Park, Kim, & Choi, 2018) and blood donors detection (Kauten, Gupta, Qin, & Richey, 2021). Compared to other applications, rumor classification using ensemble learning techniques has been very little studied.

Kaur, Kumar, and Kumaraguru (2020) propose a multilevel voting model for the fake news detection task. The study concludes that the proposed model outperforms both individual machine learning and ensemble learning models. To address the multiclass fake news detection problem, Kaliyar, Goswami, and Narang (2019) use gradient boosting ensemble techniques and compare their performance with several individual machine learning models. Results demonstrate the effectiveness of the ensemble framework compared to existing benchmark performance. Finally, Al-Ash, Putri, Mursanto, and Bustamam (2019) find that the bagging approach provides superior performance than Support Vector Machines (SVMs), Multinomial Naïve Bayes (MNB) and Random Forest to detect fake news.

3 MONITOR

Microblog messages contain rich multimodal resources, such as text contents, surrounding social context and attached images. Our focus is to leverage this multimodal information to determine whether a message is true or false. Based on this idea, we propose a framework for verifying the veracity of messages. MONITOR's detailed description is presented in this section.

3.1 Multimodal Fusion Overview

Figure 2 shows a general overview of MONITOR, which works in two main stages. First, we extract several features from the message's text and the social context. Then, we apply a feature selection algorithm to identify relevant features, which form a first set of textual features. From the attached image, we derive statistics and efficient visual features inspired from the IQA field, which form a second set of image features. Second, we train a model by concatenating and normalizing the textual and image features sets to form a fusion vector. Several machine learning classifiers may learn from the fusion vector to distinguish the veracity of the message, i.e., real or fake.

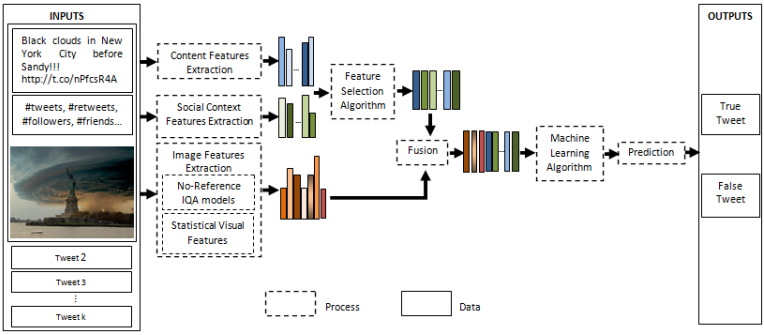


Fig. 2 Overview of MONITOR

3.2 Feature Extraction and Selection

To better extract features, we reviewed the best practices followed by information professionals, e.g., journalists, in verifying content generated by social network users. We based our thinking on relevant data from journalistic studies (Martin & Comm, 2014) and the Verification Handbook (Silverman, 2014). We define a set of features that are important to extract discriminating characteristics of rumors. These features are mainly derived from three principal aspects of news information: content, social context and visual content. The feature selection process is only applied to content and social context features sets to remove the irrelevant features that can negatively impact performance. Because our focus is the visual features set, we retain all these features in the learning process.

3.2.1 Message Content Features

Content features are extracted from the message's text. We extract characteristics such as the length of a tweet and the number of words. We also include statistics such as the number of exclamation and question marks, as well as binary features indicating the existence or not of emoticons. Furthermore, other features are extracted from the linguistics of a text, including the number of positive and negative sentiment words. Additional binary features indicate whether the text contains personal pronouns.

We also calculate a readability score for each message using the Flesch Reading Ease method (Kincaid, Fishburne Jr, Rogers, & Chissom, 1975). The higher this score is, the easier the text is to read. Other features are extracted from the informative content provided by the specific communication style of the Twitter platform, such as the number of retweets, mentions (@), hashtags (#) and URLs.

3.2.2 Social Context Features

The social context reflects the relationships between different users. Therefore, social context features are extracted from the behavior of users and the propagation network. We capture several features from the users' profiles, such as the number of followers and friends, the number of tweets the user has authored, the number of tweets the user has liked and whether the user is verified by the social media. We also extract features from the propagation tree that can be built from tweets and retweets, such as the depth of the retweet tree. Tables 1 and 2 describe the sets of content features and social context features extracted from each message.

To improve the performance of MONITOR, we apply a feature selection algorithm on the feature sets listed in Tables 1 and 2. The details of the feature selection process are discussed in Section 4.

Table 1 Content features

Description
of chars, words
of (?), (!) mark
of uppercase chars
of positive, negative words
of mentions, hashtags, URLs
of happy, sad mood emoticon
of 1st, 2nd, 3rd order pronoun
Readability score

Table 2 Social context features

Description
of followers, friends, posts
Friends/followers ratio, times listed
of retweets, likes
The user shares a homepage URL
The user has a profile image
The user has a verified account
of tweets the user has liked

3.2.3 Image Features

To differentiate between false and real images in messages, we propose to exploit visual content features and visual statistical features that are extracted from the joined images.

Visual Content Features

Usually, a news consumer decides the image veracity based on his subjective perception, but how do we quantitatively represent the human perception of the quality of an image? The quality of an image means the amount of visual degradations of all types present in an image, such as noise, blocking artifacts, blurring, fading and so on.

The IQA field aims to quantify human perception of image quality by providing an objective score of image degradations based on computational models (Maître, 2017). Such degradations are introduced during different processing stages, such as image acquisition, compression, storage, transmission and decompression. Inspired by the potential relevance of IQA metrics in our context, we use these metrics in an original way, for a purpose different from what they were created for. More precisely, we hypothesize that the quantitative evaluation of the quality of an image can be useful for veracity detection.

IQA is mainly divided into two areas of research: full-reference evaluation and no-reference evaluation. Full-reference algorithms compare the input image against a pristine reference image with no distortion. In no-reference algorithms, the only input is the image whose quality is to be measured. In our case, we do not have the original version of the posted image. Therefore, the approach that is fitting to our context is no-reference evaluation. We use three no-reference algorithms that have been demonstrated to be highly efficient: the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) by Mittal, Moorthy, and Bovik (2011), the Naturalness Image Quality Evaluator (NIQE) by Mittal, Soundararajan, and Bovik (2012) and the Perception based Image Quality Evaluator (PIQE) by Venkatanath, Praneeth, Bh, Channappayya, and Medasani (2015).

For example, Figure 3 displays the BRISQUE score computed for a natural image and its distorted versions (compression, noise and blurring distortions). The BRISQUE score is a non-negative scalar in the range [1, 100]. Lower values of the score reflect a better perceptual image quality.

Original image 13.7215 | JPEG compressed 22.6603 | Gaussian noise 28.5840 | Median blur 41.5620

Fig. 3 BRISQUE score computed for a natural image and its distorted versions

No-reference IQA metrics are also good indicators for other types of image modifications, such as GAN-generated images. These techniques allow modifying the context and semantics of images in a very realistic way. Unlike many image analysis tasks, where both reference and reconstructed images are available, images generated by GANs may not have any reference image. This is the main reason for using no-reference IQA for evaluating this type of fake images. Figure 4 displays the BRISQUE score computed for real and fake images generated by image-to-image translation based on GANs (Zhu, Park, Isola, & Efros, 2017).

Real image 17.7778 | Fake image 22.0260 | Real image 12.5000 | Fake image 22.5279

Fig. 4 BRISQUE score computed for real and fake GANs images

Statistical Features

From attached images, we define four statistical features from two aspects.

- *Number of images:* A user can post one, several or no images. To denote this feature, we count the total number of images in a rumor event and the ratio of posts containing more then one image.
- *Spreading of images:* During an event, some images are very replied and generate more comments than others. The ratio of such images is calculated to indicate this feature. Table 3 illustrates the description of our visual and statistical features. We use all of these features in the learning process.

3.3 Model Training

So far, we have obtained a first set of relevant textual features through a feature selection process. We have also a second set of image features composed of

Table 3 Description of image features

Type	Feature	Description
Visual features	BRISQUE	BRISQUE score of a given image
	PIQE	PIQE score of a given image
	NIQE	NIQE score of a given image
Statistical features	Count_Img	Number of all images in a news event
	Ratio_Img1	Ratio of the multi-image tweets in all tweets
	Ratio_Img2	Ratio of image number to tweet number
	Ratio_Img3	Ratio of the most widespread image in all distinct images

statistical and visual features. These two sets of features are scaled, normalized and concatenated to form the multimodal representation of a given message, which is learned by a supervised classifier. Several learning algorithms can be implemented fore message veracity classification. We investigate the algorithms that provide the best performance in Section 4.

4 Regular Machine Learning Experiments

In this section, we conduct extensive experiments on two public datasets. First, we present statistics about the datasets we use. Then, we describe the experimental settings: a brief review of state-of-the-art features for news verification and a selection of the best of these textual features as baselines. Finally, we present experimental results and analyze the features to achieve insights with MONITOR.

4.1 Datasets

To evaluate MONITOR's performance, we conduct experiments on two well-established public datasets for rumor detection. The detailed statistics of these two datasets are listed in Table 4.

4.1.1 MediaEval

MediaEval (Boididou et al., 2015) is collected from Twitter and includes all three characteristics: text, social context and images. It is designed for message-level verification. The dataset has two parts: a development set containing about 9,000 rumor and 6,000 non-rumor tweets from 17 rumor-related events; a test set containing about 2,000 tweets from another batch of 35 rumor-related events. We remove tweets without any text nor image, thus obtaining a final dataset including 411 distinct images associated with 6,225 real and 7,558 fake tweets, respectively.

4.1.2 FakeNewsNet

FakeNewsNet (Shu, Mahudeswaran, et al., 2018) is one of the most comprehensive fake news detection benchmark. Fake and real news articles are collected from the fact-checking websites PolitiFact and GossipCop. Since we are particularly interested in images in this work, we extract and exploit the image

information of all tweets. To keep the dataset balanced, we randomly choose 2,566 real and 2,587 fake news events. After removing tweets without images, we obtain 56,369 tweets and 59,838 images.

Table 4 MediaEval and FakeNewsNet statistics

Dataset	Set	Tweets		Images
		Real	Fake	
MediaEval	Training set	5,008	6,841	361
	Testing set	1,217	717	50
FakeNewsNet	Training set	25,673	19,422	47,870
	Testing set	6,466	4,808	11,968

4.2 Experimental Settings

4.2.1 Baseline Features

We compare the effectiveness of our feature set with the best textual features from the literature. First, we adopt the 15 best features extracted by Castillo et al. (2011) to analyze the information credibility of news propagated through Twitter. We also collect a total of 40 additional textual features from the literature (A. Gupta et al., 2013; M. Gupta, Zhao, & Han, 2012; Kwon et al., 2013; K. Wu et al., 2015), which are extracted from text content, user information and propagation properties (Table 5).

Table 5 Features from the literature

Feature
Fraction of (?), (!) Mark, # of messages
Average # of words, char lengths
Fraction of 1st, 2nd, 3rd pronouns
Fraction of URLs, @, #
Count of distinct URLs, @, #
Fraction of popular URLs, @, #
The tweet includes pictures
Average sentiment score
Fraction of positive and negative tweets
of distinct people, loc, org
Fraction of people, loc, org
Fraction of popular people, loc, org
of Users, fraction of popular users
of followers, followees, posted tweets
The user has a Facebook link
Fraction of verified users, org
of comments on the original message
Time between original message and repost

Table 6 Best textual features selected

MediaEval	FakeNewsNet
Tweet_Length	Tweet_Length
Num_Negwords	Num_Words
Num_Mentions	Num_Questmark
Num_URLs	Num_Upperchars
Num_Words	Num_Exclmark
Num_Upperchars	Num_Hashtags
Num_Hashtags	Num_Negwords
Num_Exclmark	Num_Poswords
Num_Thirdpron	Num_Followers
Times_Listed	Num_Friends
Num_Tweets	Num_Favorites
Num_Friends	Times_Listed
Num_Retweets	Num_Likes
Has_Url	Num_Retweets
Num_Followers	Num_Tweets

Table 7 Hyper-parameters configuration space

Model	Main hyper-parameters	Type	Search space
CART	max_depth	Discrete	[1,21]
	criterion	Categorical	['gini','entropy']
KNN	n_neighbors	Discrete	[1,21]
SVM	C	Discrete	[0.1,2.0]
	γ (RBF kernel)	Discrete	[0.1,1.0]
	Kernel	Categorical	['linear', 'poly', 'rbf','sigmoid']
RF	n_estimators	Discrete	[10,500]
	max_depth	Discrete	[3,20]

4.2.2 Feature Sets

The features labeled *Textual* are the best features selected among message content and social context features (Tables 1 and 2). We select them with the information gain ratio method (Karegowda, Manjunath, & Jayaram, 2010), which helps select a subset of 15 relevant textual features with an information gain larger than zero (Table 6).

The features labeled *Image* are all the image features listed in Table 3. The features labeled *MONITOR* are the feature set that we propose, consisting of the fusion of textual and image feature sets. The features labeled *Castillo* are the above-mentioned best 15 textual features. Eventually, the features labeled *Wu* are the 40 textual features identified in literature.

4.2.3 Model Construction

We cannot know beforehand what model will be good for our problem or what configuration to use. By analyzing both datasets, we found that classes are partially linearly separable in some dimensions. Thus, we evaluate a mix of simple linear and non-linear algorithms. The best result are achieved by four supervised classification algorithms: Classification and Regression Trees (CART), k-Nearest Neighbors (KNN), Support Vector Machines (SVMs) and Random Forest (RF). Then, we optimize the hyper-parameters of each model (Table 7) by testing multiple settings using the *GridSearchCV* function from the Python Scikit-Learn library (Pedregosa et al., 2011). Subsequently, we perform training and validation for each model through a 5-fold cross-validation to obtain stable out-sample results. To implement the models, we again use scikit-learn. Note that, for MediaEval, we retain the same data split scheme. For FakeNewsNet, we randomly divide data into training and testing subsets with the ratio 0.8:0.2. Table 8 present the results of our experiments.

4.3 Classification Results

From the classification results recorded in Table 8, we can make the following observations.

Table 8 Performance of individual machine learning models

Model	Features	MediaEval				FakeNewsNet			
		Acc	**Prec**	**Rec**	F_1	**Acc**	**Prec**	**Rec**	F_1
CART	Textual	0.673	0.672	0.771	0.718	0.699	**0.647**	0.652	0.65
	Image	0.632	0.701	0.639	0.668	0.647	0.595	0.533	0.563
	MONITOR	**0.746**	**0.715**	**0.897**	**0.796**	**0.704**	0.623	**0.716**	**0.667**
	Castillo	0.643	0.711	0.648	0.678	0.683	0.674	0.491	0.569
	Wu	0.65	0.709	0.715	0.711	0.694	0.663	0.593	0.627
KNN	Textual	0.707	0.704	0.777	0.739	0.698	0.67	0.599	0.633
	Image	0.608	0.607	0.734	0.665	0.647	0.595	0.533	0.563
	MONITOR	**0.791**	**0.792**	**0.843**	**0.817**	**0.758**	**0.734**	**0.746**	**0.740**
	Castillo	0.652	0.698	0.665	0.681	0.681	0.651	0.566	0.606
	Wu	0.668	0.71	0.678	0.693	0.694	0.663	0.593	0.627
SVM	Textual	0.74	0.729	0.834	0.779	0.658	0.657	0.44	0.528
	Image	0.693	0.69	0.775	0.73	0.595	0.618	0.125	0.208
	MONITOR	**0.794**	**0.767**	**0.881**	**0.82**	**0.771**	**0.743**	**0.742**	**0.743**
	Castillo	0.702	0.761	0.716	0.737	0.629	0.687	0.259	0.377
	Wu	0.725	0.763	0.73	0.746	0.642	0.625	0.394	0.484
RF	Textual	0.747	0.717	0.879	0.789	0.778	0.726	0.768	0.747
	Image	0.652	0.646	0.771	0.703	0.652	0.646	0.771	0.703
	MONITOR	**0.962**	**0.965**	**0.966**	**0.965**	**0.889**	**0.914**	**0.864**	**0.889**
	Castillo	0.702	0.727	0.723	0.725	0.714	0.669	0.67	0.67
	Wu	0.728	0.752	0.748	0.75	0.736	0.699	0.682	0.691

4.3.1 Performance Comparison

With MONITOR, using both image and textual feature allows all classification algorithms to achieve better performance than baselines. Among the four classification models, RF generates the best accuracy: 96.2% on MediaEval and 88.9% on FakeNewsNet, performing 26% and 18% better than Castillo and 24% and 15% than Wu, still on MediaEval and FakeNewsNet, respectively.

Compared to the 15 "best" textual feature set, RF improves the accuracy by more than 22% and 10% with image features only. Similarly, the other three algorithms achieve an accuracy gain between 5% and 9% on MediaEval and between 5% and 6% on FakeNewsNet. Eventually, all classification algorithms generate a lower accuracy when using image features only.

While image features play a crucial role in rumor verification, we must not ignore the effectiveness of textual features. The role of image and textual features is complementary. When the two sets of features are combined, performance is significantly boosted.

4.3.2 Illustration by Example

To more clearly show the complementarity between text and images, we compare the results achieved with MONITOR and single modality approaches (text only or image only). Fake rumor messages from Figure 1 (Section 1) are

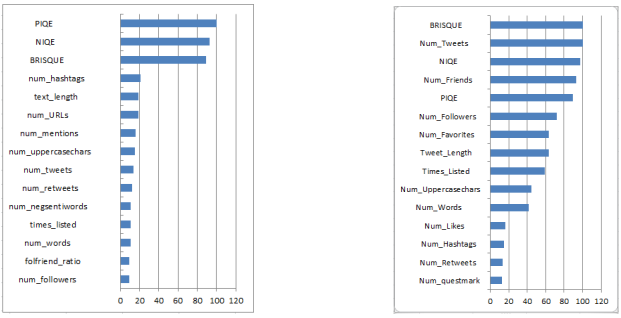
correctly detected as false by MONITOR, while using either only textual or only image modalities yields a true result.

In the tweet from Figure 1a, the text content solely describes the attached image without giving any signs about the veracity of the tweet. This is why the textual modality identifies this tweet as real. It is the attached image that looks quite suspicious. By combining textual and image contents, MONITOR can identify the veracity of the tweet with a high score, exploiting some clues from the image to get the right classification.

The tweet from Figure 1b is an example of rumor correctly classified by MONITOR, but incorrectly classified when only using the visual modality. The image seems normal and its complex semantics are very difficult to capture by the image modality. However, the words with strong emotions in the text indicate that it might be a suspicious message. By combining the textual and image modalities, MONITOR can classify the tweet with a high confidence score.

4.4 Feature Analysis

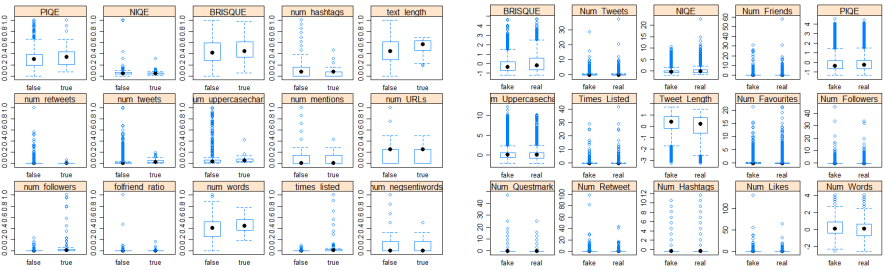
The advantage of our approach is that we can achieve some elements of interpretability. To this aim, we conduct an analysis to illustrate the importance of each feature set. We depict the first most 15 important features achieved by RF in Figure 5, which shows that, for both datasets, visual characteristics are in the top-five features. The remaining features are a mix of text content and social context features. These results validate the effectiveness of the IQA image features, as well as the the importance of fusing several modalities in the process of rumor verification.



(a) MediaEval (b) FakeNewsNet

Fig. 5 Random Forest feature importance

Eventually, to illustrate the discriminating capacity of these features, we deploy box plots for each of the 15 top variables on both datasets. Figure 6 shows that several features exhibit a significant difference between fake and real classes, which explains our good results.



(a) MediaEval (b) FakeNewsNet

Fig. 6 Distribution of true and false classes for top-15 important features

4.5 Early and Late Fusion

In our previous experiments, we fuse visual and textual modalities into a single multimodal vector before the learning and classification steps, in the so-called *early fusion* manner. Another way to merge features is *late fusion*.

This class of fusion scheme works at the decision level, by combining the prediction scores available for each modality. Late fusion starts with the extraction of unimodal features. In contrast to early fusion, where features are combined into a multimodal representation, late fusion approaches learn directly from unimodal features. The predicted probability scores are combined afterwards to yield a final detection score. Several methods help combine scores, such as averaging, voting or using another machine learning method to learn how to best combine predictions.

To apply late fusion, we train two Random Forest (RF) classifiers by learning separately the visual and textual features (Figure 7).

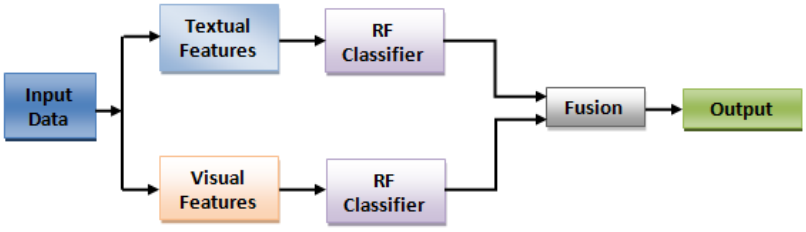
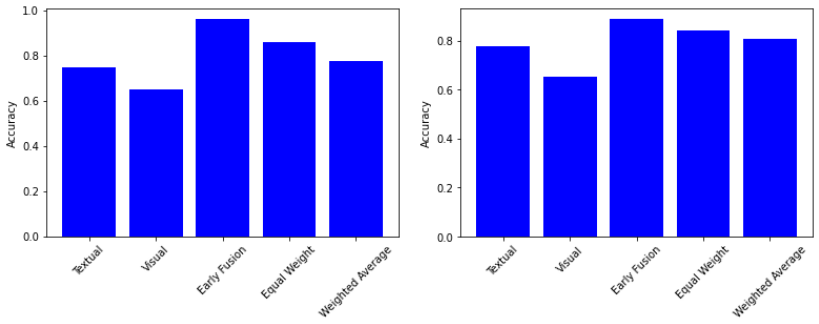


Fig. 7 Late fusion scheme

To obtain the final classification results, the predicted probabilities of the both classifiers are combined with (1) equal weights, by assuming that the two models are equally skillful and make the same proportional contribution to the final prediction; and (2) averaging the (optimized) weights by feeding the classifiers' output to a logistic regression model.

Figure 8 shows that, for both datasets, the early fusion method and the two late fusion strategies, i.e., equal weight and optimized weight, boost the prediction with different rates using separately two sets of features. Early fusion has the highest performance score, while for both late fusion techniques, equal weight is slightly more efficient than optimized weight.

Late fusion's performance is lower than that of early fusion because, when we train two models separately on visual and textual features, some dependencies between features are lost. Practically, there are some correlations between features, e.g., between BRISQUE and Num_Mention or between PIQE and Text_Length. The potential loss of correlation in the mixed feature space is a drawback of late fusion. Another disadvantage of late fusion is its cost in terms of learning effort, as every modality requires a separate supervised learning stage. Moreover, the combined representation requires an additional learning stage.



(a) MediaEval (b) FakeNewsNet

Fig. 8 Performance of early and late fusion

5 Ensemble Learning Performance

Applied machine learning often involves fitting and evaluating models on a dataset. Given that we cannot know what model will perform best on the dataset beforehand, this may involve a lot of trial and error until we find a model that performs good enough. This is akin to making a decision using the single expert we can find. A complementary approach is to prepare multiple, different models, and then combine their predictions using an ensemble machine learning model.

Because ensemble learning strategies such as bagging and boosting typically involve a single machine learning algorithm (generally a decision tree), we use instead the stacking strategy (also called metalearning) that seeks for a diverse group of members by varying model types. Figure 9 summarizes the key elements of a stacking ensemble:

- an unchanged training dataset;
- various machine learning algorithms (base models) for each ensemble member;
- a machine learning model (metamodel) to learn how to best combine predictions.

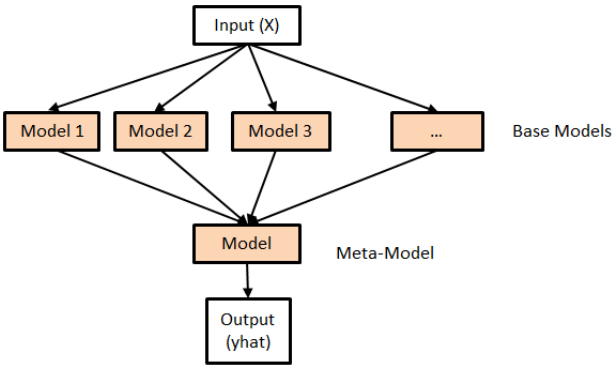


Fig. 9 Stacking ensemble

To measure the performance of ensemble learning models for rumor detection, we develop five metamodels as variants of the stacking strategy.

5.1 Metamodels

5.1.1 Voting Ensemble

We construct two voting models. The first one is a soft voting model called MONITOR$_{sv}$ that sums the predictions made by the classification models listed in Table 8 and predicts the class label with the largest sum probability. The second model is a weighted average voting model called MONITOR$_{wav}$ where model votes are proportional to model performance. The performance of each ensemble model on the training dataset will be used as the relative weighting of the model when making predictions. Performance is calculated using classification accuracy as a ratio of correct predictions ranging between 0 and 1, with larger values meaning a better model and, in turn, more contribution to the prediction.

5.1.2 Canonical Stacking Ensemble

Following Wolpert (1992)'s canonical stacking strategy (Figure 8), we construct a model called MONITOR$_{st}$. Concretely, we use three repeats of a stratified 10-fold cross-validation on the four classification models to prepare the training dataset (predictions) with the logistic regression metamodel. Furthermore, we train the metamodel on the prepared dataset as well as the

original training dataset using a 5-fold cross-validation. This aims to provide an additional context to the metamodel to better combine predictions.

5.1.3 Blending ensemble

Blending was the term commonly used for stacking ensembles during the Netflix prize in 2009. The prize involved teams seeking movie recommendations that performed better than the native Netflix algorithm. A one million US dollar prize was awarded to the team achieving a 10% performance improvement.

In this stacking-type ensemble, base models are fit on the training dataset and the metamodel is trained on predictions made by each base model on the validation dataset. At the time we are writing this paper, Scikit-learn does not support blending. Thus, we implement a blending model called MONITOR$_{bld}$ using scikit-learn models.

To implement our model, we need to split the dataset, first into training and test sets. Then, the training set is split again into two subsets used to train base models and the metamodel, respectively. We use a 50/50 split on the training and test sets and a 67/33 split on the train and validation sets (Figure 10). Furthermore, we choose logistic regression as a metamodel (the blender), for the same reasons we mentioned about canonical stacking. We summarise the key implementation steps of our model in Algorithm 1.

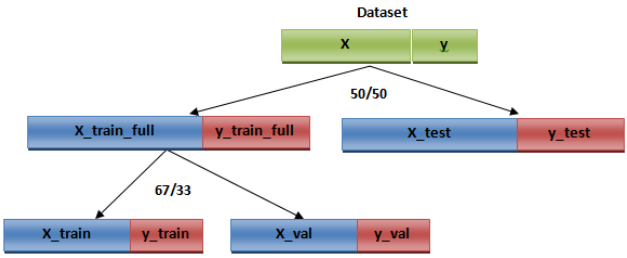


Fig. 10 Dataset splitting

5.1.4 Super Learner Ensemble

A super learner ensemble (Van der Laan, Polley, & Hubbard, 2007) is a specific stacking configuration where all base models use the same k-fold splits of data, and a metamodel is fit on the out-of-fold predictions from each model. We summarize this procedure in Algorithm 2. Moreover, Figure 11, which is reproduced from the original paper by (Van der Laan et al., 2007), depicts its data flow. We use the MLENS Python library (Flennerhag, 2017) to implement the super learner model called MONITOR$_{sl}$, where we split the training data into $k = 10$ folds. The number of base models is set to $m = 4$(i.e. KNN, CART, SVM and RF).

Algorithm 1 Blending Ensemble

Require: *Dataset*(`X,y`) ▷ input variables and output label

`meta_x, meta_y` ← `empty list`
Split *Dataset* into `X_train, y_train`, `X_val, y_val` and `X_test, y_test` ▷ train, validation and test sets
Create base models
▷ Fit the blending ensemble
for all base-model **do**
    Fit base-model on training set (`X_train, y_train`)
    Predict with base-model on `X_val`
    Store predictions in `meta_x`
end for
Convert `meta_x` to 2D array ▷ as input for blending model
Define blending model
Fit blending model on predictions from base models (`meta_x,y_val`)
▷ Make prediction with blending ensemble
for all base-model **do**
    Predict with base-model on `X_test`
    Store predictions in `meta_y`
end for
Convert `meta_y` to 2D array ▷ as input for blending model
Predict with blending model on `meta_y`
Evaluate blending model on `y_test`

Algorithm 2 Super learner ensemble

Select a k-fold split of the training dataset
Select m base-models or model configurations
for all base-model **do**
    Evaluate using k-fold cross-validation
    Store all out-of-fold predictions
    Fit the model on the full training dataset and store
end for
Fit a metamodel on the out-of-fold predictions
Evaluate the model on a holdout dataset or use model to make predictions

Table 9 summarizes the results achieved by the best individual machine learning model (RF) and the five stacking algorithms.

5.2 Result Analysis

Our comparative analysis of experimental results shows that all metalearning models are more efficient than the best individual machine learning model (RF), because by combining multiple models, the errors from a single base-model are likely compensated by the other models. As a result, the overall

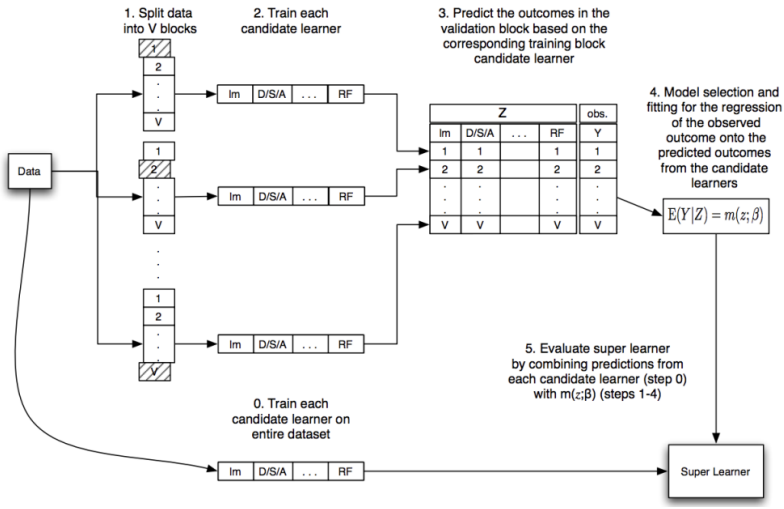


Fig. 11 Super learner ensemble data flow (Van der Laan et al., 2007)

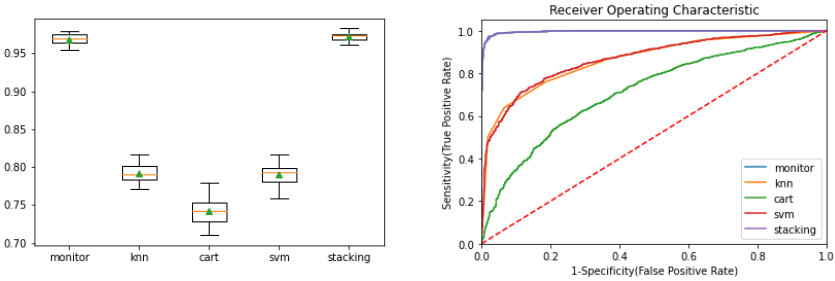
Table 9 Performance of MONITOR and stacking ensemble models

Model	MediaEval				FakeNewsNet			
	Acc	**Prec**	**Rec**	F_1	**Acc**	**Prec**	**Rec**	F_1
MONITOR	0.962	0.965	0.966	0.965	0.889	0.914	0.864	0.889
MONITOR$_{sv}$	0.966	0.955	0.976	0.965	0.897	0.911	0.873	0.892
MONITOR$_{wav}$	0.968	0.968	0.970	0.969	0.906	0.90	0.927	0.914
MONITOR$_{st}$	**0.984**	**0.979**	**0.989**	**0.984**	**0.936**	**0.929**	**0.952**	**0.941**
MONITOR$_{bld}$	0.973	0.975	0.971	0.973	0.915	0.909	0.932	0.921
MONITOR$_{sl}$	0.970	0.980	0.959	0.969	0.921	0.915	0.937	0.926

prediction performance of the ensemble is better than that of any single base-model.

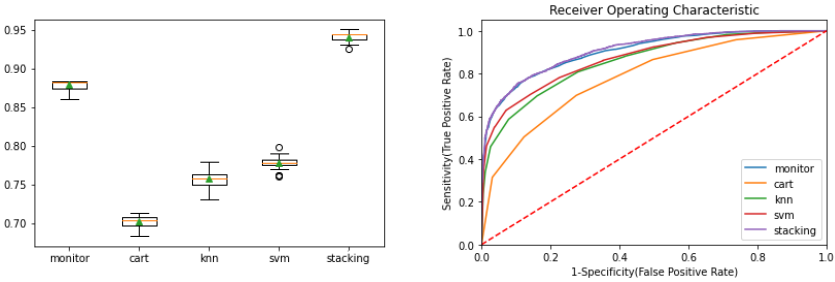
Moreover, for both datasets, the canonical stacking algorithm outperforms all models with 98.4% and 93.6% of accuracy on MediaEval and FakeNewsNet dataset, respectively. The stacking model indeed takes advantages from the diversity of predictions made by contributing models. That is, all algorithms are skillful on the classification problem, but in different ways. Figures 12 and 13 depicts the accuracy score box plot and the Receiver Operating Curve (ROC) for the canonical stacking ensemble model compared to the standalone machine learning algorithms (MONITOR-RF, CART, KNN and SVM) on MediaEval and FakeNewsNet, respectively.

Among the five ensemble models, the soft voting algorithm achieves the worst results, because it treats all models the same, i.e., all models contribute equally to the prediction. Although the canonical stacking algorithm performs the best, the blending and super learner algorithms achieve scores that are very close to those of stacking and therefore turn to be useful too for rumor classification.



(a) Accuracy (b) ROC

Fig. 12 Stacking ensemble model vs. standalone models on MediaEval



(a) Accuracy (b) ROC

Fig. 13 Stacking ensemble model vs. standalone models on FakeNewsNet

6 Conclusion and Perspectives

To assess the veracity of messages posted on social networks, most of the existing techniques ignore visual contents and use traditional machine learning models for classification, although ensemble approaches are considered the state-of-the-art solutions for many machine learning challenges. Thence, in this paper, to improve the performance of message verification, we propose a multimodal fusion framework called MONITOR that uses features extracted from the textual content of messages, the social context and image features that have not been considered until now. We compare the performance of MONITOR with five metalearning ensemble models by combining four base-predictors (KNN, CART, SVM and RF). Extensive experiments conducted on the MediaEval benchmark and the FakeNewsNet dataset show that:

- the image features that we introduce play a key role in message veracity assessment;
- no single homogeneous feature set can generate the best results alone;

- all ensemble algorithms outperform the best single base-model (RF), and canonical stacking achieves the best performance on both datasets.

Our future research includes two directions. In the short term, we plan to experiment with other, larger datasets and vary the type, combination and number of base models in the ensemble. Second, we plan to compare MONITOR's performance with a deep learning-based approach for rumor classification, deepMONITOR (Azri, Favre, Harbi, Darmont, & Noûs, 2021a), with the aim of studying the tradeoff between classification accuracy, computing complexity and explainability.

7 Declarations

Conflict of Interest

The authors have no conficting interests to declare that are relevant to the content of this article.

References

Al-Ash, H.S., Putri, M.F., Mursanto, P., Bustamam, A. (2019). Ensemble learning approach on indonesian fake news classification. *3rd International Conference on Informatics and Computational Sciences 2019 (ICICoS)* (pp. 1–6).

Azri, A., Favre, C., Harbi, N., Darmont, J., Noûs, C. (2021a). Calling to CNN-LSTM for Rumor Detection: A Deep Multi-channel Model for Message Veracity Classification in Microblogs. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases(ECML PKDD 2021)* (Vol. 12979, pp. 497–513). Bilbao, Spain.

Azri, A., Favre, C., Harbi, N., Darmont, J., Noûs, C. (2021b). MONITOR: A Multimodal Fusion Framework to Assess Message Veracity in Social Networks. *25th European Conference on Advances in Databases and Information Systems(ADBIS 2021)* (Vol. 12843, pp. 73–87).

Blei, D.M., Ng, A.Y., Jordan, M.I. (2003). Latent dirichlet allocation. *the Journal of Machine Learning Research*, *3*, 993–1022.

Boididou, C., Andreadou, K., Papadopoulos, S., Dang-Nguyen, D.-T., Boato, G., Riegler, M., ... others (2015). Verifying multimedia use at mediaeval 2015. *MediaEval*, *3*(3), 7.

Boididou, C., Papadopoulos, S., Zampoglou, M., Apostolidis, L., Papadopoulou, O., Kompatsiaris, Y. (2018). Detection and visualization

of misleading content on Twitter. *International Journal of Multimedia Information Retrieval*, *7*(1), 71–86.

Castillo, C., Mendoza, M., Poblete, B. (2011). Information credibility on twitter. *Proceedings of the 20th international conference on World wide web* (pp. 675–684).

Flennerhag, S. (2017, November). *Ml-ensemble.* Retrieved from https://dx.doi.org/10.5281/zenodo.1042144 10.5281/zenodo.1042144

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems* (Vol. 27, pp. 2672–2680).

Gupta, A., Lamba, H., Kumaraguru, P., Joshi, A. (2013). Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. *Proceedings of the 22nd international conference on World Wide Web* (pp. 729–736).

Gupta, D., & Rani, R. (2020). Improving malware detection using big data and ensemble learning. *Computers & Electrical Engineering*, *86*, 106729.

Gupta, M., Zhao, P., Han, J. (2012). Evaluating event credibility on twitter. *Proceedings of the 2012 SIAM International Conference on Data Mining* (pp. 153–164).

Gutierrez-Espinoza, L., Abri, F., Namin, A.S., Jones, K.S., Sears, D.R. (2020). Fake reviews detection through ensemble learning. *arXiv preprint arXiv:2006.07912*.

Jin, Z., Cao, J., Zhang, Y., Zhou, J., Tian, Q. (2017). Novel visual and statistical image features for microblogs news verification. *IEEE transactions on multimedia*, *19*(3), 598–608.

Kaliyar, R.K., Goswami, A., Narang, P. (2019). Multiclass fake news detection using ensemble machine learning. *IEEE 9th International Conference on Advanced Computing (IACC)* (pp. 103–107).

Karegowda, A.G., Manjunath, A., Jayaram, M. (2010). Comparative study of attribute selection using gain ratio and correlation based feature selection. *International Journal of Information Technology and Knowledge Management*, *2*(2), 271–277.

Kaur, S., Kumar, P., Kumaraguru, P. (2020). Automating fake news detection system using multi-level voting model. *Soft Computing*, *24*(12), 9049–9069.

Kauten, C., Gupta, A., Qin, X., Richey, G. (2021). Predicting blood donors using machine learning techniques. *Information Systems Frontiers*, 1–16.

Kincaid, J.P., Fishburne Jr, R.P., Rogers, R.L., Chissom, B.S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel.

Kwon, S., Cha, M., Jung, K. (2017). Rumor detection over varying time windows. *PloS one*, *12*(1), e0168344.

Kwon, S., Cha, M., Jung, K., Chen, W., Wang, Y. (2013). Prominent features of rumor propagation in online social media. *2013 IEEE 13th international conference on data mining* (pp. 1103–1108).

Lee, J., Wang, W., Harrou, F., Sun, Y. (2020). Reliable solar irradiance prediction using ensemble learning-based models: A comparative study. *Energy Conversion and Management*, *208*, 112582.

Lee, K., Park, J., Kim, I., Choi, Y. (2018). Predicting movie success with machine learning techniques: ways to im-prove accuracy. *Information Systems Frontiers*, *20*(3), 577–588.

Li, J., Li, X., Yang, B., Sun, X. (2014). Segmentation-based image copy-move forgery detection scheme. *IEEE transactions on information forensics and security*, *10*(3), 507–518.

Maître, H. (2017). *From photon to pixel: the digital camera handbook*. John Wiley & Sons.

Martin, N., & Comm, B. (2014). Information verification in the age of digital journalism. *Special Libraries Association (SLA)* (pp. 8–10).

Mittal, A., Moorthy, A.K., Bovik, A.C. (2011). Blind/referenceless image spatial quality evaluator. *2011 conference record of the forty fifth asilomar conference on signals, systems and computers (ASILOMAR)* (pp. 723–727).

Mittal, A., Soundararajan, R., Bovik, A.C. (2012). Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, *20*(3), 209–212.

Morris, M.R., Counts, S., Roseway, A., Hoff, A., Schwarz, J. (2012). Tweeting is believing?: understanding microblog credibility perceptions. *Proceedings of the acm 2012 conference on computer supported cooperative work* (pp. 441–450).

Pang, Y., Xue, X., Namin, A.S. (2016). Early identification of vulnerable software components via ensemble learning. *15th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 476–481).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825–2830.

Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R. (2018, August). Automatic detection of fake news. *Proceedings of the 27th international conference on computational linguistics* (pp. 3391–3401). Santa Fe, New Mexico, USA: ACL. Retrieved from https://www.aclweb.org/anthology/C18-1287

Pham, K., Kim, D., Park, S., Choi, H. (2021). Ensemble learning-based classification models for slope stability analysis. *Catena*, *196*, 104886.

Ruchansky, N., Seo, S., Liu, Y. (2017). Csi: A hybrid deep model for fake news detection. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797–806).

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1249.

Sangamnerkar, S., Srinivasan, R., Christhuraj, M., Sukumaran, R. (2020). An ensemble technique to detect fabricated news article using machine learning and natural language processing techniques. *International*

Conference for Emerging Technology (INCET) (pp. 1–7).

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H. (2018). Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. *arXiv preprint arXiv:1809.01286*.

Shu, K., Wang, S., Liu, H. (2018). Understanding user profiles on social media for fake news detection. *IEEE Conference on Multimedia Information Processing and Retrieval* (pp. 430–435).

Silverman, C. (2014). *Verification handbook: An ultimate guideline on digital age sourcing for emergency coverage*. European Journalism Centre.

Singh, P.D., Kaur, R., Singh, K.D., Dhiman, G. (2021). A novel ensemble-based classifier for detecting the covid-19 disease for infected patients. *Information Systems Frontiers*, *23*(6), 1385–1401.

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Van der Laan, M.J., Polley, E.C., Hubbard, A.E. (2007). Super learner. *Statistical applications in genetics and molecular biology*, *6*(1).

Venkatanath, N., Praneeth, D., Bh, M.C., Channappayya, S.S., Medasani, S.S. (2015). Blind image quality evaluation using perception based features. *Twenty First National Conference on Communications (NCC)* (pp. 1–6).

Volkova, S., & Jang, J.Y. (2018). Misleading or falsification: Inferring deceptive strategies and types in online news and social media. *Companion Proceedings of the The Web Conference 2018* (pp. 575–583).

Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., ... Gao, J. (2018). Eann: Event adversarial neural networks for multi-modal fake news detection. *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining* (pp. 849–857).

Wolpert, D.H. (1992). Stacked generalization. *Neural networks*, *5*(2), 241–259.

Wu, K., Yang, S., Zhu, K.Q. (2015). False rumors detection on sina weibo by propagation structures. *IEEE 31st international conference on data engineering* (pp. 651–662).

Wu, L., & Liu, H. (2018). Tracing fake-news footprints: Characterizing social media messages by how they propagate. *Proceedings of the eleventh ACM international conference on Web Search and Data Mining* (pp. 637–645).

Zhang, C., & Ma, Y. (2012). *Ensemble machine learning: methods and applications*. Springer.

Zhu, J.-Y., Park, T., Isola, P., Efros, A.A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE international conference on computer vision.*

Authors' Biographies

Abderrazek Azri received the Master's degree in information systems security from the University of Lyon 2, Lyon, France, in 2014, he received the Ph.D. degree from the University of Lyon 2, Lyon, France, in july 2022. His research interests include social media analysis and data mining.

Cécile Favre is a lecturer in computer science at the University of Lyon 2 and a member of the ERIC laboratory. She is also an associate researcher at the Centre Max Weber. His work focuses on the one hand on computer science research (decision-making computing, social network analysis with in particular work in the field of bibliometrics). On the other hand, she is developing interdisciplinary work around gender issues and IT. She is also involved in various responsibilities in the Mention of Master in Gender Studies in Lyon.

Nouria Harbi is a member of research staff ERIC Laboratory, University Lyon 2, France. She received her master degree in computer science and Ph.D. from INSA Lyon, France. She then joined the laboratory ISEOR, where she worked on information systems. She is currently working on the security of decisional information system and modelling data warehouse.

Jérôme Darmont is full professor of computer science at the University of Lyon, France. He received his Ph.D. in 1999 from the University of Clermont-Ferrand II, France, and then joined the University Lyon 2 as an associate professor. He became full professor in 2008 and has been director of the ERIC research center from 2012 to 2021. In 2017, he was made Honoris Causa professor at Simon Kuznets Kharkiv National University of Economics, Ukraine. He is currently adjunct director of the Institute of Communication at University Lyon 2. His research interests mainly relate to data management performance (performance optimization, auto-administration and benchmarking of databases, data warehouses, data lakes, data meshes...) and cloud business intelligence (data security, query performance and cost, personal BI, big data analytics, textual document analysis...).

Camille Nôus came into existence on 20 March 2020, to represent the contribution of the academic community to research in France, in the form of a collective and gender-neutral signature. This signature, devised as a scientific consortium, calls for an open and collaborative approach to the creation and diffusion of knowledge, under the aegis of the academic community, and is intended to be a mark of integrity.