Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

An Intelligent Framework for Automated Fake News Detection in Social Networks Using GRU-LSTM Deep Neural Network

FIRST A. AUTHOR[1], (Fellow, IEEE), SECOND B. AUTHOR[2], AND THIRD C. AUTHOR, JR.[3], (Member, IEEE)

[1]National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: author@boulder.nist.gov)
[2]Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu)
[3]Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA

Corresponding author: First A. Author (e-mail: author@ boulder.nist.gov).

This paragraph of the first footnote will contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

ABSTRACT Social media has become an essential part of our life. It has become one of the most important sources of information. Due to the harmful impact of fake news on society, it becomes essential to develop efficient computerized fake news detection systems. In this work, we propose a Deep Neural Network for fake news detection. The deep learning techniques are LSTM (one to three layers) and GRU (one to three layers). We compare the performance of the proposed approaches with Six machine learning techniques. The six machine learning techniques are decision tree (DT), logistic regression (LR), K nearest neighbor (KNN), random forest (RF), support vector machine (SVM), and naive Bayes (NB). The parameters of deep learning techniques are optimized using Keras-tuner, while the parameters of machine learning techniques are optimized using a grid search. Three Benchmark datasets were used to train and test models. Two feature extraction methods were used (TF-ID with N-gram) to extract essential features from the three benchmark datasets for the baseline machine learning model and word embedding feature extraction method for deep neural network methods. The proposed deep learning techniques always show the best performance because of their ability to learn the discriminatory features through the multiple hidden layers.LSTM(one layer) showed the best cross-validation accuracy (82.68%) on the first dataset. In the case of the second dataset, LSTM(two layers) showed the best cross-validation accuracy (94.21%). Finally, in the third dataset, GRU (two layers) showed the best cross-validation accuracy (86.05%). These results demonstrate significant improvement in the area of fake news detection as compared to the existing state of art results of baseline machine learning models.

INDEX TERMS Fake News, social media, Deep Learning, Machine Learning, Optimization .

I. INTRODUCTION

Fake news is one type of many misinformation types, such as rumors and false content on the web. Many studies [1] [2] have defined fake news as "news articles that are intentionally written to mislead or misinform readers, but can be verified as false by means of other sources". Fake news has been distinguished for contributing to expanded political polarization and fanatic strife in later times. Recent examples included the contention made amid the 2016 presidential campaign for the US [3] and Indian Airstrike in Balakot in 2019. Distinguishing between genuine news and fake news is required by building AI systems. It is a challenging errand for social media stages such as Facebook, Twitter, etc. to recognize the fake contents inside the massive volume of information posted by the users. Fake news can be categorized into three categories serious fabrications, large scale hoaxes, and humorous fakes [4]. The serious fabrications category includes faked interviews and malicious intent articles that spread through social media. Large scale hoaxes are fake information that appears as if they are proper news. The last category is humorous fakes, which are satirical news to entertain users, but they are disguised as real news. The credibility of the web sites that publish news is the main factor that determines the correctness of news. Many malicious web sites mimic the domain name of trustful websites. Such websites can be used to harvest fake news to form a dataset for analysis. However, fake news can be found on trustful websites just by mistake. To form a dataset, both truthful and fake news articles have to be collected. It takes human experts to detect the veracity of the news. Depending on human expert verification is not the only way to validate news. Crowdsourcing and computational oriented fact checking models are used as alternatives to annotate the fake news [5]. A benchmark dataset of fake news has not been agreed until now because of the trouble of collecting fake news and the ambiguity in providing a clear definition [5] .However, some authors created a dataset from the statements collected from social media such as LIAR [6]. Some authors altered Wikipedia sentences to produce statements and provide evidence for or against such claims in Wikipedia articles [7]. Another fake news dataset is collected from Facebook and Twitter [8] [9] [10] [11]. Finally, the complete dataset is provided in [12], where the authors provided a dataset that contains information about the content and the social context of the news. Researches on fake news make a difference between content features and context features. The content features of fake news are linguistic features. The context features of fake news are the surrounding information such as user's characteristics and social network-based features [5].

Hence, there exists a considerable risk of publishing fake over social media. Therefore, the truth of news is the need of the hour, and this research is the move towards addressing this critical problem. figure 1 shows some of the examples of fake news spread over Social Media. Recently, an example was seen in France and the USA. figure 1.a and 1.b show fake news about the Presidential candidates during the France and US presidential election, which was shared over thousands of times, and spread quickly. figure 1.a to figure 1.e showed examples of fake news spread quickly and how content broadcasts over social media, whereas figure 1.f showed examples of the adverse effects of fake news on people and the decision toward the results of fake news. In figure 1.a [13], The first example of Fake news, Titled by Emmanuel Macron"s presidential campaign, is financed by Saudi Arabia. It is published on 24 February 2017 and claimed that the campaign of Emmanuel Macron, a centrist candidate, was financially supported by Saudi Arabia (30% of Macron's campaign funded by Saudi Arabia during France presidential election.). The facts showed that The story is fake. The site appeared on - LeSoir.info - copied the design and layout of the real Le Soir website - LeSoir.be - to spread false information. The story URL has generated almost 10,000 likes, shares, and comments on Facebook. In figure 1.b [13], The second example of Fake news, Titled by Michele Obama, deletes Hillary Clinton from twitter. This fake news causes adverse effects on the outcome of the US presidential elections. In figure 1.c [13], The third example of fake news (A tweet shared over 1,700 times), titled by Marine Le Pen, criticized the "Masha and the Bear" cartoon because the little girl in the story wears a veil. It is published on 26 February 2017 and claimed that French presidential candidate and National Front leader Marine Le Pen criticized a children"s cartoon, Masha and the Bear, because Masha wore a "veil." The facts showed that Marine Le Pen did not tweet that - a screenshot of the image is doctored. Satirical website secretnews.fr published a 2014 article on the same topic. In figure 1.d [13], The fourth example of Fake news, Titled by The French state, is replacing Christian public holidays with Muslim and Jewish holidays. It is published on 12 March 2017 and claimed that two new Muslim and Jewish public holidays would be introduced. The facts showed that The story is fake, and The government has never announced such a proposal, although a think tank, Terra Nova, has advocated the idea. The story URL has generated approximately 6,000 likes, shares, and comments on Facebook. In figure 1.e [14], The fifth example of Fake news, Titled by The French state, will spend (100 million euros) buying hotels on housing migrants. It is published on 10 March 2017 and claimed that the Council of Europe Development Bank (CEB) would lend the French state 100 million euros (87.3 million dollars) to purchase hotels for housing asylum seekers. The facts showed that The story is fake and Two separate news stories have been conflated and altered, distorting the truth and exaggerating the assistance available to asylum seekers. The story URL has generated close to 10,000 likes, shares, and comments on Facebook. Finally, in figure 1.f show a real tweet by US President Trump .he has ordered a funding freeze on the World Health Organization(WHO), accusing it of practicing "blackout" over the spread of the Coronavirus. He considered that the WHO made "mistakes" during the emerging Coronavirus (Covid-19) crisis, which caused the death of thousands. He said that The World Health Organization had provided false information obtained from China, which has spread the virus on its territory. "He also considered that the international organization" was late in warning about the danger of the Coronavirus. It failed to obtain sufficient information about it and publish it transparently. More and more, bogus news stories on Facebook and other social media have painted politicians in a false light. Some recent examples have claimed - falsely - to have turned up examples of bad behavior by political candidates or public officials. Others portray politicians as heroes for deeds they never did. Either way, fake news erodes trust in the news media. Moreover, tall tales can sway public opinion.

According to the previous examples leads to the problem of the user being misinformed. The cumulative effect of misinformed users on social media has a very negative impact. The spreading of false information also hampers the public emotionally. We tried to solve this problem with our proposed model. The paper contributions of the proposed techniques can be summarized as follows:

1) We proposed a novel Fake News Detection system on social media platforms using A Deep Neural Network.
2) Improving the accuracy of existing fake news detection using Deep Neural Network

3) We optimized the results of the Deep Learning Algorithms with hyper-parameters optimization methods.
4) We perform a comparison based study with other states of the art machine learning algorithms for ensuring the results of the proposed deep learning algorithms
5) We are applying different feature selection methods such as n-gram sizes with TF-IDF and word embedding.
6) The proposed system Achieved 94.21% accuracy in Fake News Detection on the best dataset.

The remainder of this paper is organized as follows: the related work is presented in section 2. The proposed methodology is introduced in section 3. The experiment results and discussion are discussed in section 4 and section 5, respectively. Finally, conclusions are presented in section 6.

II. RELATED WORK

Detecting fake news becomes one of the most critical tasks of artificial intelligence scientists. There are two main approaches for detection: the machine learning approach and the Deep Learning approach.

1) **Machine learning approach for fake news detection** : Deception has been studied and defined as the creation of a false conclusion by transmitting a false message. In their study, [15] analyzed a set of linguistic features and investigated three classifiers. SVM achieved the highest precision, recall, and F-measure. However, linguistic features and visual features are commonly used in SVM approaches such as [15] [16] [17] [18] [19]. In [20] , authors distinguished between fake news, satire news, and real ones by introducing a set of distinguishing features such as the titles. In [21] , both content and context-based features were used to detect fake news using a Decision Tree(DT). However, Random Forest (RF) and Decision Tree (DT) were applied using user characteristics in [22] [23] to detect the trustworthiness of users writing the news. Additional features like topic models based features are used in [21] . In [17] , the authors defined some linguistic cues of deception and applied Random Forest (RF) to detect fake news. Logistic Regression (LR) has shown competitive performance in detecting fake news in [24] [10] [25].
2) **Deep learning approach for fake news detection:** Deep learning classifiers have become popular in recent years. The approach of deep learning is efficient in terms of extracting relevant features [26] . Recurrent neural networks (RRNs) and, in particular, LSTM is efficient in modeling sequential data [27] . In [28] , the authors proposed different RNN architectures, namely tanh-RNN, LSTM, and Gated Recurrent Unit (GRU), and GRU achieved the best performance. In [29] , LSTM has been fed by a mix of content and context-based features of news, and it achieved good accuracy in detecting fake news. CNN's are a class of neural networks that gain popularity in the NLP field [30] . In [31] , both RNN and CNN are used to detect false news and show a better performance than the performance of baselines. In [32] , the authors used LSTM and hybrid LSTM-CNN architectures. The simplest LSTM showed the best performance. In [6] , a hybrid model of RNNs and CNNs was used by the authors where the text information is encoded via CNN, and LSTM encodes the metadata of the author. This hybrid model outperformed the baseline model. 1 represents a summary of related work

III. METHODOLOGY

The proposed system of fake news detection consists of two main categories, as shown in Figure 2. The first category uses regular machine learning algorithms, and the second category by using deep neural networks. The first category detects fake news using six baseline traditional machine learning techniques. The machine learning techniques are decision tree (DT), logistic regression (LR), K-nearest neighbor (KNN), random forest (RF), support vector machine (SVM), and Naive Bayes (NB). The second category detects fake news using two deep learning techniques. The two deep learning techniques are LSTM (one to three layers) and GRU (one to three layers). **In the first step (preprocessing step)** for the two categories; includes removing unimportant characters, tokenization, removing stop wording, and stemming. **In the third step**, the feature is extracted using TF-IDF with N-grams for the first category, ML techniques. In contrast, for the second category, i.e., deep learning techniques, the feature is extracted using the word embedding method with Glove to build a word embedding matrix. **In the fourth step**, the parameters of traditional machine learning techniques are optimized using a grid search with stratified cross-validation, while the parameters of deep learning techniques are optimized using a Keras-tuner library. **The performance of each technique** is evaluated by measuring accuracy, precision, recall, and F-Measure. Each step is described in detail in the following subsections.

A. DATA COLLECTION

Experiments were conducted using three Twitter fake news datasets in different topics, including disasters [33], PolitiFact [34], gossip cop [34]. The disaster dataset is collected from Kaggle about the topic disaster, while the second and third datasets are related PolitiFact and gossip cop topics collected from FakeNewsNet.

1) The disaster dataset has five features; id, text, location, keyword, and target (see Table 2). The Disaster dataset has a text of 7613 tweets. Each given tweet is about a real disaster or not labeled as 1 and 0, respectively. In particular, 4342 tweets show a real disaster, while 3271 shows not. In our experiment, we have used two features, which are text and target, as a label, which shows whether a tweet is about a real disaster (1) or not (0).

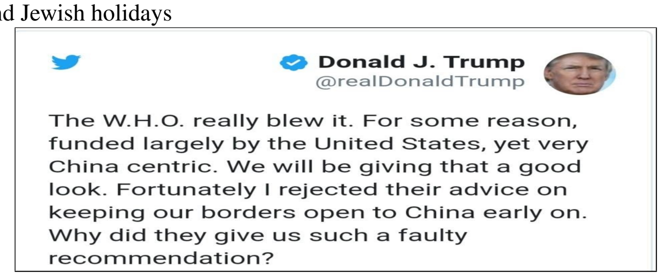
(a) Emmanuel Macron's presidential campaign is financed by Saudi Arabia

(b) Michele Obama deletes Hillary Clinton from twitter

(c) Marine Le Pen criticized the "Masha and the Bear" cartoon because the little girl in the story wears a veil

(d) The French state is replacing Christian public holidays with Muslim and Jewish holidays

(e) The French state is going to spend 100 million euros buying hotels on housing migrants



(f) The US President threatens to stop funding WHO because its data is incorrect

Figure 1: Example of some fake news over social media and some reactions on the fake news

Table 1: Comparison between different Fake News frameworks

Ref#	Authors	Title	method	features	Year
[25]	L. Zhou, D.P. Twitchell, T. Qin, J.K. Burgoon, J.F. Nunamaker	An exploratory study into deception detection in text-based computer-mediated communication	Logistic regression	content	2003
[21]	C. Castillo , M. Mendoza , B. Poblete	Information credibility on twitter	Decision tree	Content and context	2011
[15]	H. Zhang, Z. Fan, J.-h. Zheng, Q. Liu	An improving deception detection method in computer-mediated communication	SVM	Content-based features	2012
[16]	S. Afroz, M. Brennan, R. Greenstadt	Detecting hoaxes, frauds, and deception in writing style online	SVM	Content-based features	2012
[28]	K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio	Learning phrase representations using RNN encoder-decoder for statistical machine translation	Deep learning	Content and context	2014
[17]	E.J. Briscoe, D.S. Appling, H. Hayes	Cues to deception in social media communications	SVM	Content-based features	2014
[23]	J. Ito, J. Song, H. Toda, Y. Koike, S. Oyama	Assessment of tweet credibility with LDA features	Random forest	Content and context	2015
[18]	V. Perez-Rosas , R. Mihalcea	Experiments in open domain deception detection,	SVM	Content-based features	2015
[24]	M. Hardalov, I. Koychev, P. Nakov	In search of credible news	Logistic regression	content	2016
[19]	V. Rubin, N. Conroy, Y. Chen, S. Cornwell	Fake news or truth? Using satirical cues to detect potentially misleading news	SVM	Content-based features	2016
[20]	B.D. Horne, S. Adali	This just in: fake news packs a lot in the title, uses more straightforward, repetitive content in the text body, more similar to satire than real news	SVM	Content-based features	2017
[10]	E. Tacchini, G. Ballarin, M.L. Della Vedova, S. Moret, L. de Alfaro	Some like it hoax: automated fake news detection in social networks	Logistic regression	content	2017
[6]	W.Y. Wang	Liar, liar pants on fire: a new benchmark dataset for fake news detection	ensemble	Content and context	2017
[31]	S. Volkova, K. Shaffer, J.Y. Jang, N. Hodas	Separating facts from fiction: linguistic models to classify suspiciously and trusted news posts on twitter	RNN and CNN	Content and context	2017
[29]	N. Ruchansky, S. Seo, Y. Liu, CSI	a hybrid deep model for fake news detection	LSTM	Content and context	2017
[6]	W.Y. Wang	Liar, liar pants on fire: a new benchmark dataset for fake news detection	Deep learning	Content and context	2017
[32]	O. Ajao, D. Bhowmik, S. Zargari	Fake news identification on twitter with hybrid CNN and RNN models	RNN and CNN	Content and context	2018

Table 2: The disaster dataset description

Features	Description
Id	a unique identifier for each tweet
text	the text of the tweet
location	the location the tweet was sent from (may be blank)
keyword	a particular keyword from the tweet (may be blank)
target	target denotes whether a tweet is about a real disaster (1) or not (0)

Table 3: The final PolitiFact and final gossip cop datasets description

Features	Description
Id	Unique identifiers for each news
URL	Url of the article from the web that published that news
title	Title of the news article
tweet_ids	Tweet ids of tweets sharing the news. This field is a list of tweet ids separated by a tab.
label	The label denotes whether a tweet is about a real disaster (1) or not (0)

2) The PolitiFact dataset has two files 1) politifact_real.csv, which contains samples related to real news that includes 432 tweets, 2) politifact_fake.csv contains samples related to fake news that includes 618 tweets. We merged politifact_real.csv and politifact_fake.csv files into one file where each tweet belongs to politifact_real labeled as 0 while each tweet belongs to politifact_fake labeled as 1. The final PolitiFact dataset has five features: id, URL, title, tweet-id, and label (3). In our experiment, we used the title to represent the text of the tweet and label features.
3) The gossip cop dataset has two files 1) gossip cop_real.csv, which contains sample tweets related to real news that includes 5328 tweets, 2) gos-

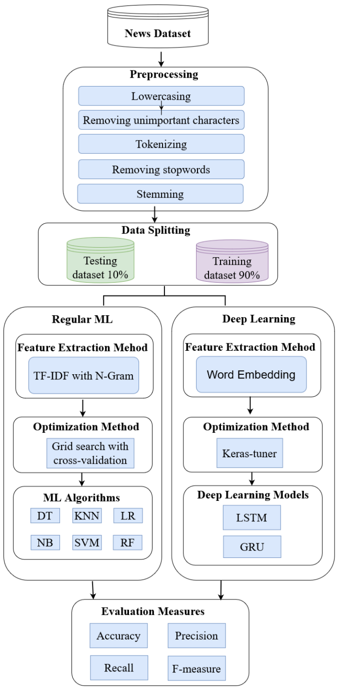


Figure 2: The proposed system of fake news detection.

sipcop_fake.csv contains sample tweets related to fake news. We selected 5322 tweets from gossipcop_fake.csv. We merged gossipcop_real.csv and gossipcop_fake.csv files into one file where each tweet belongs to gossipcop_real labeled as 0 while each tweet belongs to gossipcop_fake labeled as 1. The final gossip cop dataset has five features, including id, URL, title, tweet-id, and label (3). In our experiment, we used the title to represent the text of the tweet and label features.

B. DATA PREPROCESSING

Data preprocessing is an important phase for any sentiment analysis system, especially for social media content. Twitter data is the popular unstructured datasets collected of information from people entered his/her feelings, opinion, attitudes, products review, emotions, etc. These datasets need to be subjected to certain refinements by performing preprocessing techniques to the next phases of elaboration, i.e., applying ML/DL techniques. The basic cleaning operations within preprocessing techniques used in this work are removing unimportant characters, stop-word removal, tokenization, a lower casing, sentence segmentation, and punctuation removal. They will help us to reduce the size of actual data by removing the irrelevant information that exists in the data and then to achieve better performances. In our study, the preprocessing involves a series of techniques which are listed in the following steps:

- **Lower casing:** simply, it one of the basic cleaning operations to convert a word to lower cases such as NLP -> nlp.
- **Removing unimportant data:** the punctuation like commas, apostrophes, quotes, question marks, and more which do not add much value to a natural language model are deleted.
- **Tokenization:** It is the key aspect of working with text data to separate a piece of text into smaller units called tokens. The tokens are including paragraphs and sentences which can be further broken into words. For example, consider this sentence before tokenization: "never give up", after tokenization it comes 'never', 'give,' 'up'.
- **Removal of Stop Word:** a stop word usually refers to the most common words in a language that does not add much meaning to a sentence such as articles, prepositions and conjunctions, and some pronouns. These words are removed from each tweet with the datasets.
- **Stemming:** stemming is removing the suffix from and transform it to its root word to reduce the number of word types or classes in the data. For example, the words "Making," "Made," and "Maker" will be reduced to the word "make."

C. DATA SPLITTING

In this step, datasets are split into 90% of the training dataset and 10% of the testing dataset using a stratified method. The training set is fed into the ML/DL models to let our models learn from this data, while the unseen test set is used to evaluate ML/DL models.

D. APPLYING OPTIMIZATION AND LEARNING MODELS

In this step, we applied using two Learning models (Regular Machine Learning algorithms, Deep Learning models). Further details about each model are presented as follows:

1) Regular Machine Learning

We applied six machine learning models after two steps .firstly is by applying Machine Learning Feature Extraction Method then optimized the models using Hyperparameters Optimization Methods. Further details about each model are presented as follows:

1) **Machine Learning Feature Extraction Method :**In this step, we have used the TF-IDF feature extraction method with different sizes of N-Gram and 3000 matrix size. N-gram is the simplest model that assigns probabilities to sentences and sequences of words with length n. The value of n can be 1, 2, 3, and so on, called uni-gram, bi-gram, and tri-gram. N-grams can be divided into two categories: 1) character-based and 2) word-based. A character N-gram is a set of n consecutive characters extracted from a word. The primary motivation behind this approach is that similar words will have a high proportion of N-grams in common. Typical values for n are 2 or 3; these correspond to bigrams or trigrams, respectively. For example, the word computer results in:

- the generation of the bigrams *C, CO, OM, MP, PU, UT, TE, ER, R*
- o and the generation of the trigrams **C, * CO, COM, OMP, MPU, PUT, UTE, TER, ER*, R** Where '*' denotes a padding space. Character-based N-grams are generally used in measuring the similarity of character strings. The Term Frequency-Inverted Document Frequency (TF-IDF) is a well-known feature method to evaluate the importance of a word in a document. According to this work, the TF-IDF method is used to measure the importance of a term within a tweet in the fake news datasets. The key idea of the TF-IDF method is converting the tweets into a Vector Space Model (VSM) and then calculating the importance of the term by counting its frequencies within the tweets. The word-based frequency is counted using different n-grams, including unigram, bi-gram, and tri-gram, etc.

2) **Hyperparameters Optimization Methods:** In this step, we have used the hyperparameters optimization techniques to select the best value for each parameter of regular machine learning models, including Grid Search with stratified 10-fold cross-validation described as follows:
   - **Grid search** is a hyperparameter optimization technique for hyperparameter tuning, which is used to methodically select the best value that achieves the best performances for an ML model. It evaluates ML model for each combination of algorithm parameters specified in a grid and then reports the optimal values of model hyperparameters.
   - **K-Fold Cross-Validation** is mainly used for hyperparameter tuning by dividing the sample of datasets into a training set to train the model, and a test set to evaluate it. The dataset is split into k equal partitions where k-1 groups are used for training, and the one fold is held for the testing model. This process is repeated k times (i.e., k=10), including one fold, is used for testing and k-1 folds for the training set. In our experiment, we used k = 10. In the 10-fold CV process, 90% of data were used for the training, and 10% of data were used for testing purposes.
3) **Machine Learning Models:** We have used six regular machine learning algorithms, which are Decision Tree, Random forest, K-Nearest Neighbor, Logistic Regression, Support Vector Machine, and Naive Bayes, to classify news into fake and real news. Further details about each model are presented as follows:
   - **Decision Tree (DT)** [35] are useful supervised Machine learning algorithms that can perform our classification tasks in this paper. It consists of nodes and branches, where the tests on each attribute are represented at the nodes, the outcome of this procedure is represented at the branches, and the class labels are represented at the leaf nodes. The goal is to create a model that classifying the value of a target variable by learning simple decision rules concluded from the data features.
   - **Random forest (RF)** [36], [37] is a supervised machine learning algorithm that uses a collection of decision trees, providing more flexibility, accuracy, and ease of access. This algorithm dominates over decision trees algorithm as decision trees provide low accuracy compared to the random forest algorithm. In simple words, the random forest approach increases the performance of decision trees. It is one of the best algorithms in classification techniques, and we used it in our paper. The goal is to create a model that classifying the value of a target variable by learning simple decision rules concluded from the data features from more than the tree.
   - **K-Nearest Neighbor (KNN)** [38], [39] is a Supervised classification algorithm. It is one of the most straightforward and widely used algorithms which depends on its k value; K specifies the number of neighbors, and its algorithm is as follows:
     - Choose the number K of neighbor.
     - Take the K Nearest Neighbor of a new data point, according to Euclidean Distance. (We can increase or decrease it as you like to get the best accuracy that we needed)
     - Among the K-neighbors, Count the number of data points in each category.
     - Assign the new data point to a category, where you counted the most neighbors
   - **Support vector machine (SVM)** [40], [41] are supervised learning models with associated learning algorithms that analyze data mostly used for classification problems. In this algorithm, each data item is plotted as a point in n-dimensional space (where n is the number of features), with the value of each feature being the value of a particular coordinate. Then, classification is performed by finding the hyper-plane that best differentiates the two classes. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification, implicitly mapping their inputs into high-dimensional feature spaces.
   - **Logistic Regression (LR)** [42] is a Machine Learning algorithm used for classification problems. It is a predictive analysis algorithm and based on the concept of probability. It is based on the sigmoid function where output is the probability (Value of output ranges from 0 to 1), and input can be from-infinity to +infinity. If we need to classify our data into two classes, then if the output

probability range is less than 0.5, then our data in the first class (class tag (0)), and if the probability range more than 0.5, then our data in the second class (class tag (1)).

- **Naive Bayes (NB)** [43] is a probabilistic machine learning model based on Bayes' theorem. It makes classifications using the Maximum Posterior decision rules in a Bayesian setting.

2) Deep Learning

We applied Two Deep learning models after two steps .firstly is by applying Deep Learning Feature Extraction Method then optimized the models using Hyperparameters Optimization Methods. Further details about each model are presented as follows:

1) **Deep Learning Feature Extraction Method** In this step, we have used word embeddings, which generally converts text data, i.e., words into vectors. It represents every word in an n-dimensional dense vector where similar words will have a similar vector. The more efficient word embeddings techniques which have proven there capability to convert words into vectors are GloVe and Word2Vec. According to this work, GloVe [44] represents the tweets within the fake news datasets into dense vectors, which fed into the deep learning models. The gloVe is an unsupervised learning algorithm for word embeddings, which is used to obtain the vector representations for words. The key idea of the GloVe technique is to discover the closeness of two words, with their separation in a vector space to create vector representations called word embedding vectors. The embedding vectors are created by aggregating global word-word co-occurrence statistics from the datasets and then resulting in the matrix representations, including measuring the closeness of two words in a tweet. We used glove.twitter.27B.zip that includes a different dimension of vectors, which are 25d, 50d, 100d, and 200d vectors. We used 200d vectors to build the embedding matrix.
2) **Hyperparameters Optimization Method:** For hyperparameters optimization, we have used a Keras-tuner [45] library to pick the optimal set of hyperparameters in hidden layers (LSTM or GRU) and dropout layers. We set different values for different parameters: the number of neurons, reg_rate for l2 regularization technique [46], and the dropout rate for the dropout layers [47]. For this, we have applied the Keras-tuner on the training dataset to select the best parameters, as shown in 4.
3) **Deep neural network :** 3 shows the deep neural network architecture that is used to classify news into fake and real. It consists of a word embedding matrix as input to embedding layer, embedding layer, hidden layers including Long Short-Term Memory (LSTM), or Gated Recurrent Units (GRU), flatten layer [48], and output layer. In the word embedding matrix, the GloVe word embedding technique has been used to calculate word embeddings using a co-occurrence matrix in between words within fake news tweets, which is called the embedding matrix. The embedding matrix is used to represent the tweets into dense vectors. The embedding layer, hidden layers, and output layer are described as follows:

Table 4: Hyperparameters configurations selected by Keras-tuner.

Parameter	The value
Dropout rate	within the range of 0.1 rates to 0.5 rate
The number of neurons	within the range of 10 neurons to 200 neurons
regularization l2	0.01, 0.05, 0.1, .2,.3,.4,.5

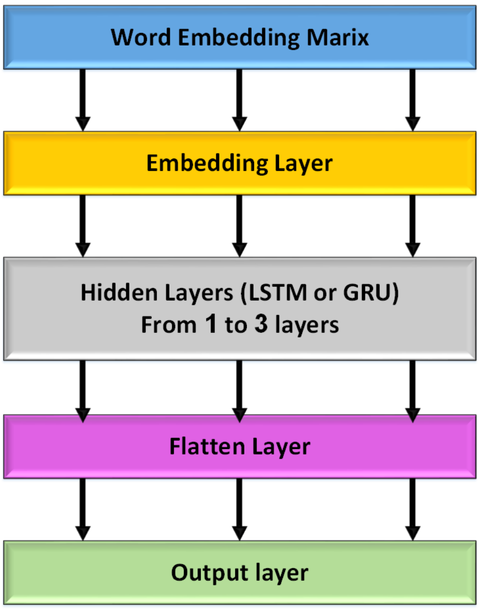


Figure 3: Deep neural network architecture.

- **Embedding layers** The embedding layer is implemented in the Keras library [?]. Regarding this work, Keras library is used to initialize the embedding layer to learn an embedding for all of the words in the training dataset. The **Keras** embedding layer has three arguments, including
  a) **input_dim** defines the size of the vocabulary in the dataset.
  b) **output_dim** defines the size of the vector space in which words will be embedded.
  c) **input_length** defines the length of input sequences as defined for any input layer of a Keras model. The embedding layer is configured as follows; **input_dim** equals 20000 because the number of words is 20000, **output_dim equals 200 and because we used 200d vectors of golvetweet and input_length equals 32.**

- **Hidden Layer** Two different neural network models are used; LSTM or GRU model. For each model, a different number of hidden layers has been applied, including one, two, and three layers. Also, one dropout layer and different numbers of neurons in each hidden layer have been used. The ReLU (Rectified Linear Unit) [**?**] activation function has been applied for the hidden layers. For each hidden layer, l2 regularization techniques have been used by adopting reg_rate value for l2. Also, we used the dropout layer and the different number of dropout rate.
- **Output Layer** The output layer provides the final output of the model where the neural network model classifies the inputs tweets into two categories; real or fake. In particular, the output layer has one neuron, which detected the news within an input tweet in terms of fake or real. In this layer, we used the ADAM optimizer [**?**] and sigmoid [**?**] is the activation function.

4) **Recurrent Neural Network (RNN) :** Not all problems can be converted into one with fixed length inputs and outputs. Problems such as Speech Recognition or Time-series Prediction require a system to store and use context Information. Hard/Impossible to choose a fixed context window. There can always be a new sample longer than anything is seen. RNNs are useful as their intermediate values (state) can store information about past inputs for a time that is not fixed a priori. An RNN is called recurrent because they perform the same task for each element in the sequence. The RNN uses the hidden state to record the state of each moment while processing the sequence data, and the current state depends on the current input and the state of the previous moment. Therefore, the current hidden state makes full use of past information [**?**]. In this way, an RNN can process sequence data in dynamic processes [**?**]. The architecture of an RNN is shown in 4. When given an input sequence $X = [x_1, x_2 \cdot \cdot x_t \cdot \cdot x_T]$ of length T, an RNN defines the hidden state ht at the time t of a sequence as in 1:

$$h_t = \tanh\left(W_h h_{t-1} + W_x x_t + b\right) \tag{1}$$

Although the RNN is very powerful when dealing with sequence problems, it is difficult to train with the gradient descent method because of the well-known gradient vanishing/explosion problem [**?**]. On the other hand, variants of RNN have been developed to solve the above problems, such as Long Short-Term Memory (LSTM), GRU, etc. Therefore, GRU is adopted in our method.

5) **Long Short-Term Memory network (LSTM):** Long Short-Term Memory network (LSTM) [**?**], [**?**] is a Deep Recurrent Neural Network (RNN) that is better than the conventional RNN on tasks involving long time lags [**?**], [**?**] .the main difference between RNN and LSTM, RNN, is a single layer (tanh) where LSTM is Four interactive layers (5). LSTM basic unit is the memory block containing one or more memory cells and three multiplicative gating units. For the gate, σ is a logistic sigmoid, ranging from 0 to 1. LSTM has three of these gates to protect and control the cell state. The three gated are the

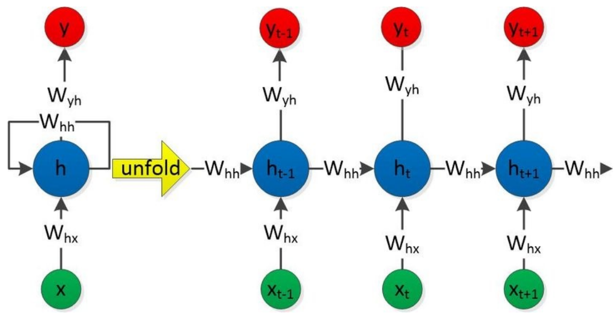


Figure 4: RNN architecture.

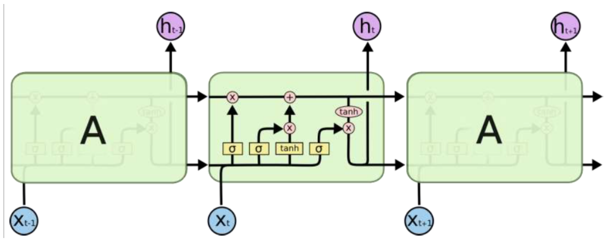


Figure 5: LSTM architecture.

- **Forget gate layer** f_t, as shown in 6. Forget gate layer is used to decide what information throw away from the cell state). Forget gate layer (2) Output a number is between 0 and 1.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2}$$

- Then **Add new information** as shown in 6 is to decide what new information store in the cell state - **Input gate** layer i_t by 3

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{3}$$

then Decides which values we will update (Tanh layer $\tilde{C}_t$) by creating a vector of new candidate values

$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{4}$$

- **Update cell state**C_t, as shown in 6: by Forgetting the things we decided to forget earlier $\mathrm{f_t} * \mathrm{C_{t-1}}$ and Adding information we decide to be added $\mathrm{i_t} * \widetilde{\mathrm{C}}_\mathrm{t}$ as shown in the following formula 5

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{5}$$

- **Create Output (Output gate layer** o_t**, Tanh layer**)by Decide what we are going to Output as shown in 6

– Output gate layer o_t as shown in 6: Decides what parts of the cell state we are going to Output

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{6}$$

– **Tanh layer**h_t: Push the values between -1 and +1

$$h_t = o_t * \tanh\left(C_t\right) \tag{7}$$

- **Peephole** 6 to Let the gate layer look at the cell state (entire/ partial) as shown in the following Equations 8

$$\begin{aligned} f_t &= \sigma\left(W_f \cdot \left[\boldsymbol{C}_{t-1}, h_{t-1}, x_t\right] + b_f\right) \\ i_t &= \sigma\left(W_i \cdot \left[\boldsymbol{C}_{t-1}, h_{t-1}, x_t\right] + b_i\right) \\ o_t &= \sigma\left(W_o \cdot \left[\boldsymbol{C}_t, h_{t-1}, x_t\right] + b_o\right) \end{aligned} \tag{8}$$

- **Coupled forgot and input gates** 6: Not deciding separately as shown in the following Equations 9

$$C_t = f_t * C_{t-1} + \left(1 - f_t\right) * \tilde{C}_t \tag{9}$$

So we can summarise the LSTM into four steps

– **Step 1:** Forget gate layer
– **Step 2:** Input gate layer
– **Step 3:** Combine step 1 and 2
– **Step 4:** Output the cell state

Although LSTM is a kind of RNN. LSTM is capable of learning long term dependencies. RNN cannot learn to connect the information in a large gap, but LSTM does not have a large gap problem. LSTM aims at minimizing the cost function. LSTM accepts variable-length character sequences such that it does not require linguistic/statistical features to be extracted [3]. This algorithm is also compact; the update complexity per weight and time step and storage complexity per weight is on the order of O (1)

- **Concept of the Gated Recurrent Unit (GRU):** GRU has the same chain structure as a simple RNN or STM, but a GRU is more complicated because it updates the hidden state. The main difference between LSTM and GRU; in GRU, Combine the forget and input layer into a single "update gate, Merge the cell state and the hidden state, and Simpler and popular [**?**], [**?**]. Instead of directly updating the current hidden state with the previous hidden state, GRU uses a reset gate and updates the gate, judging whether the information in the previous hidden state is useful, then holds useful information and removes useless information [**?**]. 7 shows the architecture of GRU. The way GRU updates h_t is as follows:

a) **The reset gate r_t and update gate z_t**

$$z_t = \sigma\left(W_{zh} h_{t-1} + W_{zx} x_t + b_z\right) \tag{10}$$

$$r_t = \sigma\left(W_{rh} h_{t-1} + W_{rx} x_t + b_r\right) \tag{11}$$

For the gate,σ is a logistic sigmoid, The reset gate rt, and update gate zt ranging from 0 to 1.

b) **Candidate hidden state $\tilde{h}_t$:**

$$\tilde{h}_t = \tanh\left(W_{\tilde{h}h}\left(r_t * h_{t-1}\right) + W_{\tilde{h}x} x_t + b_h\right) \tag{12}$$

The candidate hidden state $\tilde{h}_t$ uses the reset gate r_t to control the inflow of the previous hidden state h_{t-1} containing past information. If the reset gate is approximately zero, the previous hidden state will be removed. Therefore, the reset gate provides a mechanism to remove previous hidden states unrelated to the future; that is, the reset gate determines how much information was forgotten.

c) **hidden stateh_t:**

$$h_t = z_t * h_{t-1} + \left(1 - z_t\right) * \tilde{h}_t \tag{13}$$

The hidden state h_t uses the update gate z_t to update the previous hidden state h_{t-1} and the candidate hidden state h_t. If the update gate is approximately 1, the previous hidden state will be held and passed to the current moment. The GRU can cope with the gradient vanishing/explosion problem in the RNN, so it is more suitable for the fault diagnosis of dynamic processes.

6) **Evaluating models** Four standard metrics are used to evaluate the models, including accuracy, precision, recall, and F1-score, where TP is True Positive, TN is True Negative, FP is False Positive, and FN is a False Negative. See Equations **??**.

- **Accuracy** is a measure of totally correctly identified samples out of all the samples.

$$\text{Accuracy } = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{14}$$

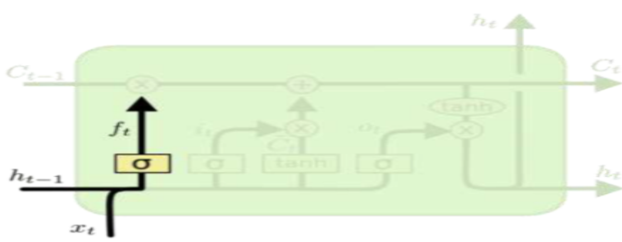
- **Precision and Recall** The measure of the ability of the model to accurately identified the occurrence of a positive class instance is determined by recall

$$\text{Precision } = \frac{TP}{TP + FP} \tag{15}$$

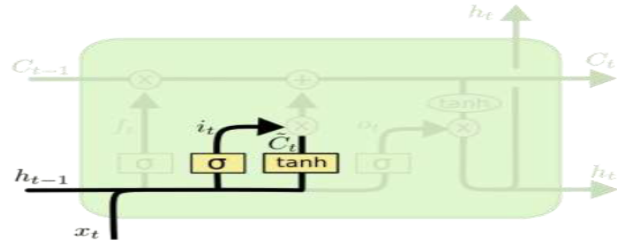
$$\text{Recall} = \frac{TP}{TP + FN} \tag{16}$$

- **F1-Score** The harmonic mean of Precision and Recall

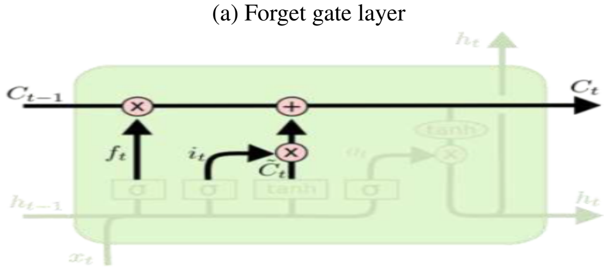
$$F1 - \text{ Score } = \frac{2 * \text{ Precision } * \text{ Recall}}{\text{Precision } + \text{ Recall}} \tag{17}$$



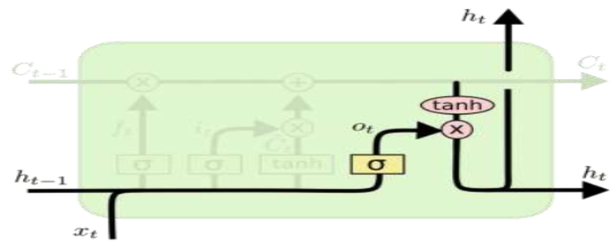
(a) Forget gate layer



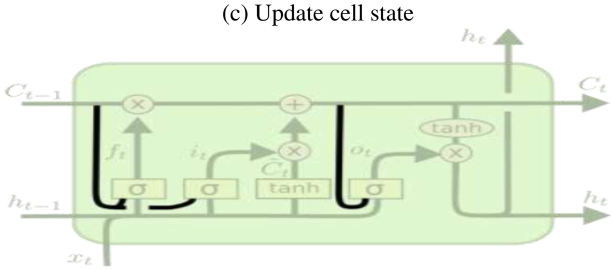
(b) Add new information (Input gate layer , Tanh layer)



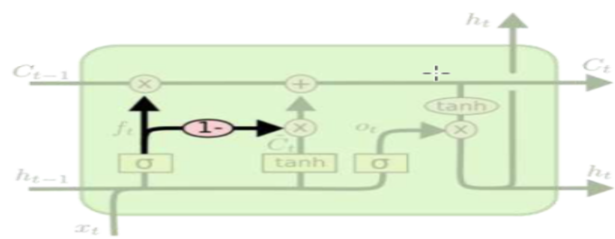
(c) Update cell state



(d) Create output (Output gate layer , Tanh layer)



(e) LSTM Peephole



(f) Coupled forgot and input gates

Figure 6: LSTM layers Steps

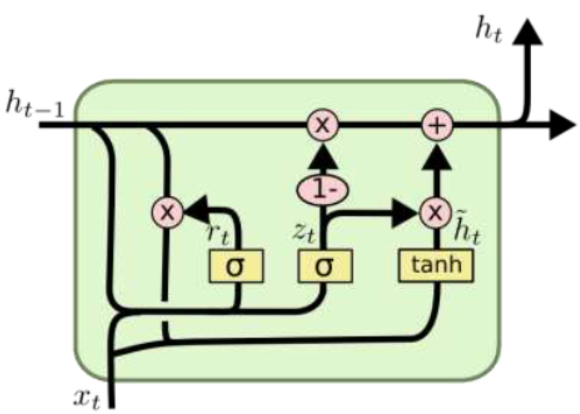


Figure 7: GRU architecture.

References

[1] G. O. Young, "Synthetic structure of industrial plastics," in Plastics, 2nd ed., vol. 3, J. Peters, Ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15–64.

[2] W.-K. Chen, Linear Networks and Systems. Belmont, CA, USA: Wadsworth, 1993, pp. 123–135.

[3] J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility," IEEE Trans. Electron Devices, vol. ED-11, no. 1, pp. 34–39, Jan. 1959, 10.1109/TED.2016.2628402.

[4] E. P. Wigner, "Theory of traveling-wave optical laser," Phys. Rev., vol. 134, pp. A635–A646, Dec. 1965.

[5] E. H. Miller, "A note on reflector arrays," IEEE Trans. Antennas Propagat., to be published.

[6] E. E. Reber, R. L. Michell, and C. J. Carter, "Oxygen absorption in the earth's atmosphere," Aerospace Corp., Los Angeles, CA, USA, Tech. Rep. TR-0200 (4230-46)-3, Nov. 1988.

[7] J. H. Davis and J. R. Cogdell, "Calibration program for the 16-foot antenna," Elect. Eng. Res. Lab., Univ. Texas, Austin, TX, USA, Tech. Memo. NGL-006-69-3, Nov. 15, 1987.

[8] Transmission Systems for Communications, 3rd ed., Western Electric Co., Winston-Salem, NC, USA, 1985, pp. 44–60.

[9] Motorola Semiconductor Data Manual, Motorola Semiconductor Products Inc., Phoenix, AZ, USA, 1989.

[10] G. O. Young, "Synthetic structure of industrial plastics," in Plastics, vol. 3, Polymers of Hexadromicon, J. Peters, Ed., 2nd ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15-64. [Online]. Available: http://www.bookref.com.

[11] The Founders' Constitution, Philip B. Kurland and Ralph Lerner, eds., Chicago, IL, USA: Univ. Chicago Press, 1987. [Online]. Available: http://press-pubs.uchicago.edu/founders/

[12] The Terahertz Wave eBook. ZOmega Terahertz Corp., 2014. [Online]. Available: http://dl.z-thz.com/eBook/zomega_ebook_pdf_1206_sr.pdf. Accessed on: May 19, 2014.

[13] Philip B. Kurland and Ralph Lerner, eds., The Founders' Constitution. Chicago, IL, USA: Univ. of Chicago Press, 1987, Accessed on: Feb. 28, 2010, [Online] Available: http://press-pubs.uchicago.edu/founders/

[14] J. S. Turner, "New directions in communications," IEEE J. Sel. Areas Commun., vol. 13, no. 1, pp. 11-23, Jan. 1995.

[15] W. P. Risk, G. S. Kino, and H. J. Shaw, "Fiber-optic frequency shifter using a surface acoustic wave incident at an oblique angle," Opt. Lett., vol. 11, no. 2, pp. 115–117, Feb. 1986.

[16] P. Kopyt et al., "Electric properties of graphene-based conductive layers from DC up to terahertz range," IEEE THz Sci. Technol., to be published. DOI: 10.1109/TTHZ.2016.2544142.

[17] PROCESS Corporation, Boston, MA, USA. Intranets: Internet technologies deployed behind the firewall for corporate productivity.

Presented at INET96 Annual Meeting. [Online]. Available: http://home.process.com/Intranets/wp2.htp
[18] R. J. Hijmans and J. van Etten, "Raster: Geographic analysis and modeling with raster data," R Package Version 2.0-12, Jan. 12, 2012. [Online]. Available: http://CRAN.R-project.org/package=raster
[19] Teralyzer. Lytera UG, Kirchhain, Germany [Online]. Available: http://www.lytera.de/Terahertz_THz_Spectroscopy.php?id=home, Accessed on: Jun. 5, 2014
[20] U.S. House. 102nd Congress, 1st Session. (1991, Jan. 11). H. Con. Res. 1, Sense of the Congress on Approval of Military Action. [Online]. Available: LEXIS Library: GENFED File: BILLS
[21] Musical toothbrush with mirror, by L.M.R. Brooks. (1992, May 19). Patent D 326 189 [Online]. Available: NEXIS Library: LEXPAT File: DES
[22] D. B. Payne and J. R. Stern, "Wavelength-switched pas- sively coupled single-mode optical network," in Proc. IOOC-ECOC, Boston, MA, USA, 1985, pp. 585–590.
[23] D. Ebehard and E. Voges, "Digital single sideband detection for interferometric sensors," presented at the 2nd Int. Conf. Optical Fiber Sensors, Stuttgart, Germany, Jan. 2-5, 1984.
[24] G. Brandli and M. Dick, "Alternating current fed power supply," U.S. Patent 4 084 217, Nov. 4, 1978.
[25] J. O. Williams, "Narrow-band analyzer," Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, USA, 1993.
[26] N. Kawasaki, "Parametric study of thermal and chemical nonequilibrium nozzle flow," M.S. thesis, Dept. Electron. Eng., Osaka Univ., Osaka, Japan, 1993.
[27] A. Harrison, private communication, May 1995.
[28] B. Smith, "An approach to graphs of linear forms," unpublished.
[29] A. Brahms, "Representation error for real numbers in binary computer arithmetic," IEEE Computer Group Repository, Paper R-67-85.
[30] IEEE Criteria for Class IE Electric Systems, IEEE Standard 308, 1969.
[31] Letter Symbols for Quantities, ANSI Standard Y10.5-1968.
[32] R. Fardel, M. Nagel, F. Nuesch, T. Lippert, and A. Wokaun, "Fabrication of organic light emitting diode pixels by laser-assisted forward transfer," Appl. Phys. Lett., vol. 91, no. 6, Aug. 2007, Art. no. 061103.
[33] J. Zhang and N. Tansu, "Optical gain and laser characteristics of InGaN quantum wells on ternary InGaN substrates," IEEE Photon. J., vol. 5, no. 2, Apr. 2013, Art. no. 2600111
[34] S. Azodolmolky et al., Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks," J. Lightw. Technol., vol. 29, no. 4, pp. 439–448, Sep. 2011.
[35] S. Azodolmolky et al., Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks," J. Lightw. Technol., vol. 29, no. 4, pp. 439–448, Sep. 2011.
[36] S. Azodolmolky et al., Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks," J. Lightw. Technol., vol. 29, no. 4, pp. 439–448, Sep. 2011.
[37] S. Azodolmolky et al., Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks," J. Lightw. Technol., vol. 29, no. 4, pp. 439–448, Sep. 2011.
[38] S. Azodolmolky et al., Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks," J. Lightw. Technol., vol. 29, no. 4, pp. 439–448, Sep. 2011.
[39] S. Azodolmolky et al., Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks," J. Lightw. Technol., vol. 29, no. 4, pp. 439–448, Sep. 2011.
[40] S. Azodolmolky et al., Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks," J. Lightw. Technol., vol. 29, no. 4, pp. 439–448, Sep. 2011.
[41]
[42]
[43]
[44]
[45]
[46]
[47]
[48]

FIRST A. AUTHOR (M'76–SM'81–F'87) and all authors may include biographies. Biographies are often not included in conference-related papers. This author became a Member (M) of IEEE in 1976, a Senior Member (SM) in 1981, and a Fellow (F) in 1987. The first paragraph may contain a place and/or date of birth (list place, then date). Next, the author's educational background is listed. The degrees should be listed with type of degree in what field, which institution, city, state, and country, and year the degree was earned. The author's major field of study should be lower-cased.

The second paragraph uses the pronoun of the person (he or she) and not the author's last name. It lists military and work experience, including summer and fellowship jobs. Job titles are capitalized. The current job must have a location; previous positions may be listed without one. Information concerning previous publications may be included. Try not to list more than three books or published articles. The format for listing publishers of a book within the biography is: title of book (publisher name, year) similar to a reference. Current and previous research interests end the paragraph. The third paragraph begins with the author's title and last name (e.g., Dr. Smith, Prof. Jones, Mr. Kajor, Ms. Hunter). List any memberships in professional societies other than the IEEE. Finally, list any awards and work for IEEE committees and publications. If a photograph is provided, it should be of good quality, and professional-looking. Following are two examples of an author's biography.

SECOND B. AUTHOR was born in Greenwich Village, New York, NY, USA in 1977. He received the B.S. and M.S. degrees in aerospace engineering from the University of Virginia, Charlottesville, in 2001 and the Ph.D. degree in mechanical engineering from Drexel University, Philadelphia, PA, in 2008.

From 2001 to 2004, he was a Research Assistant with the Princeton Plasma Physics Laboratory. Since 2009, he has been an Assistant Professor with the Mechanical Engineering Department, Texas A&M University, College Station. He is the author of three books, more than 150 articles, and more than 70 inventions. His research interests include high-pressure and high-density nonthermal plasma discharge processes and applications, microscale plasma discharges, discharges in liquids, spectroscopic diagnostics, plasma propulsion, and innovation plasma applications. He is an Associate Editor of the journal Earth, Moon, Planets, and holds two patents.

Dr. Author was a recipient of the International Association of Geomagnetism and Aeronomy Young Scientist Award for Excellence in 2008, and the IEEE Electromagnetic Compatibility Society Best Symposium Paper Award in 2011.

THIRD C. AUTHOR, JR. (M'87) received the B.S. degree in mechanical engineering from National Chung Cheng University, Chiayi, Taiwan, in 2004 and the M.S. degree in mechanical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2006. He is currently pursuing the Ph.D. degree in mechanical engineering at Texas A&M University, College Station, TX, USA.

From 2008 to 2009, he was a Research Assistant with the Institute of Physics, Academia Sinica, Tapei, Taiwan. His research interest includes the development of surface processing and biological/medical treatment techniques using nonthermal atmospheric pressure plasmas, fundamental study of plasma sources, and fabrication of micro- or nanostructured surfaces.

Mr. Author's awards and honors include the Frew Fellowship (Australian Academy of Science), the I. I. Rabi Prize (APS), the European Frequency and Time Forum Award, the Carl Zeiss Research Award, the William F. Meggers Award and the Adolph Lomb Medal (OSA).