Convolutional Neural Networks for Multimodal Fake News Detection

A RESEARCH PROJECT REPORT

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF

MASTER OF TECHNOLOGY
IN
INFORMATION SYSTEMS

Submitted by:

CHAHAT RAJ
2K19/ISY/06

Under the supervision of

Ms. PRIYANKA MEEL
ASSISTANT PROFESSOR
DEPARTMENT OF INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi college of Engineering)
Bawana Road, Delhi-110042

JULY, 2021

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi college of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Chahat Raj, Roll No. 2K19/ISY/06 student of M.Tech, Information Systems, hereby declare that the M.Tech. Research Project Report titled "Convolutional Neural Networks for Multimodal Fake News Detection" which is submitted by me to the Department of Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi
Date: July 24, 2021

Ms. Chahat Raj
(2K19/ISY/06)

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi college of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the M.Tech. Research Project Report titled “Convolutional Neural Networks for Multimodal Fake News Detection” which is submitted by Chahat Raj, Roll No. 2K19/ISY/06 Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Priyanka

Place: Delhi
Date: July 24, 2021

Ms. Priyanka Meel
SUPERVISOR
ASSISTANT PROFESSOR
DEPARTMENT OF INFORMATION TECHNOLOGY

ACKNOWLEDGEMENTS

Foremost, I am thankful to God for being able to write this in full gratitude.

I am immensely grateful to the Vice-Chancellor of Delhi Technological University, Prof. Yogesh Singh and the Head of Department of Information Technology, Prof. Kapil Sharma.

This thesis would not have its essence without the combined efforts of all the people involved.

I am immensely grateful to my mentor, **Ms. Priyanka Meel**, who has been extremely supportive since the beginning of my Masters. She is the lady who built me into a researcher. From the initial days until today, she has constantly been entailing and encouraging. Without her efforts, I would not have been what I am. Whilst I knew only what research is, she taught me how it is done. We share an untold understanding which was imperative for me to complete my thesis. Her words have been extremely motivational throughout the course of time. Thanks to all of her unfailing efforts!

To my professors, **Prof. Dinesh Kumar Vishwakarma**, **Prof. Seba Susan** and **Ms. Anamika Chauhan**; whose appreciations and expectations pushed me to keep climbing heights. To their worthy lectures and their out-of-the-classroom support. Thank you for showing confidence in me.

ABSTRACT

An upsurge of false information revolves around the internet. Social media and websites are flooded with unverified news posts. These posts are comprised of text, images, audio, and videos. There is a requirement for a system that detects fake content in multiple data modalities. We have seen a considerable amount of research on classification techniques for textual fake news detection, while frameworks dedicated to visual fake news detection are very few. We explored the state-of-the-art methods using deep networks such as CNNs and RNNs for multi-modal online information credibility analysis. They show rapid improvement in classification tasks without requiring pre-processing. To aid the ongoing research over fake news detection using CNN models, we build textual and visual modules to analyze their performances over multi-modal datasets. We exploit latent features present inside text and images using layers of convolutions. We see how well these convolutional neural networks perform classification when provided with only latent features and analyze what type of images are needed to be fed to perform efficient fake news detection. We propose a multi-modal Coupled ConvNet architecture that fuses both the data modules and efficiently classifies online news depending on its textual and visual content. We thence offer a comparative analysis of the results of all the models utilized over three datasets. The proposed architecture outperforms various state-of-the-art methods for fake news detection with considerably high accuracies.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Visual data attracts viewers more quickly than words do. A Human brain captures and rapidly analyzes a news item and often flags it as fake or real by just a glance of its title, image, or a small segment of it, mostly without going through the entire textual content. It does this based on the preexisting knowledge in our conscience. Even if it does go through a whole text, there are very few references and not enough time to check for the authenticity of the content we come across. Various content creators exploit these drawbacks of the human brain and behavior. There is a need for technological state of the art methods to assess the credibility of content, textual or visual, and authenticate it as fake or real. Online media emerged as a platform to share ideas, views, news. With the advancement of mobile devices and the internet, news became easily accessible to people who were either deprived of or uninterested in official news sources such as television and newspapers. The long and seemingly tedious to read texts became easy to understand as images and videos now accompany them. In the same process, it also became challenging to detect the truth in such content.

In the present scenario, online media is losing its charm and credibility as content creators lure users to gain popularity and money using the content they post online. In this process, they do not pay heed to the authenticity of the information, ignore the verification process, and mix up misleading or tampered images or clips with the texts. Content creators focus on posting catchy and attractive content that bags them many likes, comments, and dollars. Sometimes both the text and graphic content are intentionally made erroneous to spread fake news, making the entire content even more unrealistic. Hence there is an urgent need to design and develop a new classification method to assess the credibility of content, textual or visual, and segregate it as fake or real. If textual and visual factors are taken collectively, fake news detection methods have proved to provide higher accuracies than unimodal detection methods. Machine learning and deep learning-based detection mechanisms depend on fake or real news by analyzing the text's features and visual data. Users consuming information play an essential part in stopping the spread of fake content at the root level or circulating to reach a great mass affecting political, social, and economic lives. The algorithms so far used depend upon news data collected from websites and social media platforms, which are later classified into binary (real and fake) or multiple (ranging according to their severities) labels by crowdsourcing or third-party authenticators.

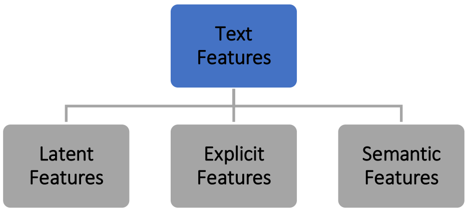

Figure 1: Text features

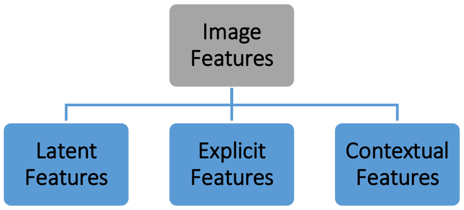

Figure 2: Image features

With the advent of massive data and news content online, the intricacies add up when multiple data forms are available. Despite being beneficial in terms of easy transmission and news consumption, multi-modal data also presents a strenuous task for detecting fake news amongst them. The modalities prevalent on online media include text, image, audio, video, and hyperlinks. With the vast accompaniment of text with visual data, the effectiveness of news rises. A large amount of visual data makes verification difficult as multi-modal data does not guarantee the credibility and attracts more attention than pure text contents. Multi-modal features are expected to be more beneficial in detecting fake news as compared to unimodal features. Few of the excellent quality datasets available for scientific research include binary labeled datasets and multi-label datasets such as Mediaeval, Sina Weibo, PolitiFact, Emergent, and Resized_V2 [1].

Figures 1 and 2 represent critical knowledge predominantly available in text and image parts of information circulating online. We propose that online social media images consist of three features: latent features, explicit features, and contextual features. Latent features are extracted using layers of convolutions. Deep convolutional networks are capable of learning kernel values that are utilized to extract latent features. According to Yang et al. [1], explicit features are hand-crafted features such as the resolution of an image and the number of faces in the picture. Apart from these two intrinsic features, contextual features are based on semantic relationships between the text and the image. We have executed convolutional neural networks for text and image classifications. CNNs provide an advantage to extract features directly from raw input without any pre-processing required. CNNs reduce input data on various layers such that only required information is preserved and worked upon to make essential predictions. In this work, we propose a novel fake news detection framework. It is based on two-stream convolutional neural networks for text and image input streams. This novel architecture consists of individual text and image classification modules, which are fused at a later stage post-training of convolutional models. The experiments performed resulted in ~3-6% higher scores than the established state-of-the-art methods. The proposed architecture is capable of detecting fake news based on both textual and visual information. The usage of Text-CNN increases the overall efficiency of the architecture. Simultaneously, the combination of Image-CNN has resulted in an additive accuracy for the detection task. The use of convolutional models that we propose with introduced Text-CNN and Image-CNN models outperform the existing state-of-the-art.

The contributions of this work include:

- Web scraping, creating clean-image datasets from two previously available datasets that contained news URLs.

- We have proposed a new Coupled ConvNet architecture that constitutes proposed Text-CNN and Image-CNN modules for multi-modal fake news detection.

- We have implemented CNN models on TI-CNN, Emergent, and MICC-F220 dataset on textual and visual data.

- We have performed a comparative analysis of various CNN models' efficiencies on real-world datasets for fake news detection.

- We have analyzed the performance of deep learning on latent textual and visual features for fake news detection.

- We have provided new deep learning pathways to better fake news detection.

CHAPTER 2

LITERATURE REVIEW

Fake news detection challenges include the usage of multi-mal data to classify real and fake news. Present methodologies include fake news detection on textual content [2-4]. Research shows that the incorporation of visual data improves fake news detection. With the rise of multi-modal content on users' posts and news contents, studies involving detection using visual data have rapidly increased.

Previous research [2,5] includes studying image features of visual data like accompanying images, type of image, etc. Other investigations include learning forensic features [6, 7]. Text information is fused by Jin et al. [8] to get better detection using attention mechanism with RNN on image and LSTM on the text and social context to obtain features and perform rumor detection on microblogs. Qi et al. [9] combined Recurrent Neural Networks to detect and interpret real and fake photos semantically. They introduced a novel approach called Multi-domain Visual Neural Network (MVNN). It uses CNN to extract frequency-domain patterns and CNN-RNN to extract pixel domain patterns and fuses using an attention mechanism outperforming state-of-the-art methods by 9.2%. Researchers have provided various forensics tools and techniques to identify image manipulations. Mostly used methods include detecting physical cues within the image.

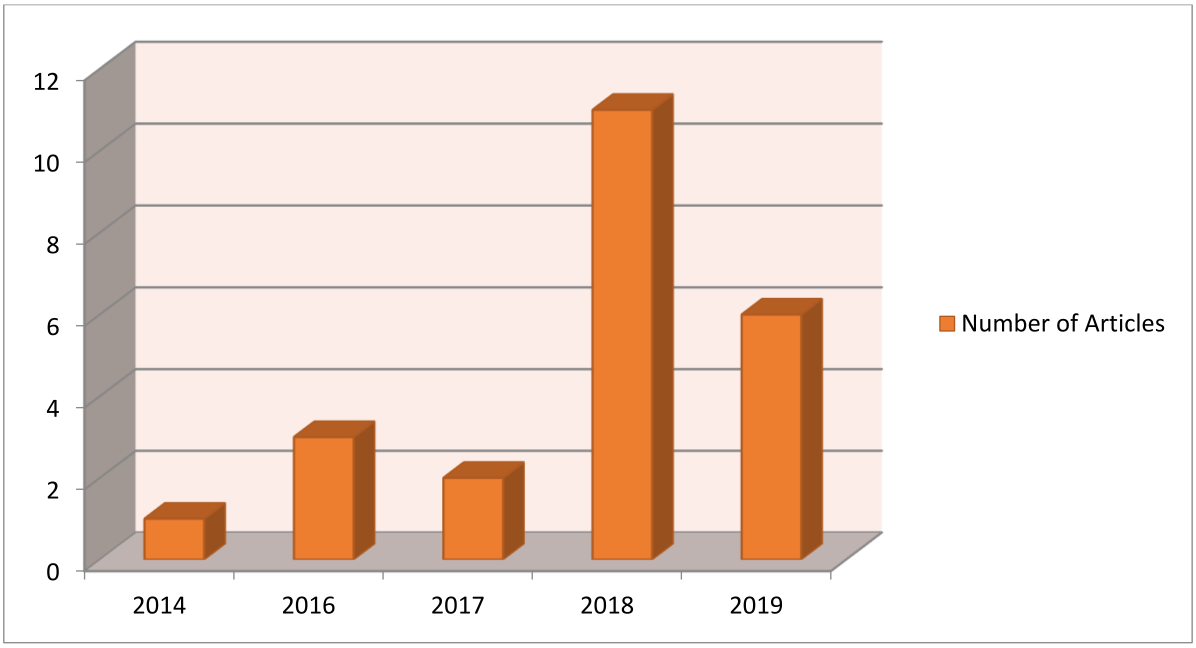

Figure 1: Yearly trend of research works

Recent works have traversed towards deep learning techniques rather than using available prior knowledge about the data. Using labeled training data is specifically advanced for fake news detection. Previous studies focused on linguistic and textual data to study fake news characteristics and semantics of the data. Deep Neural Networks have been utilized to check tweets for temporal-linguistic traits [3]. Attention mechanisms have also been used with RNNs for fusion [10]. Liu and Wu [11] modeled the classification with a combination of CNNs and RNNs. Less focus has been given to the credibility of multi-modal data on the web. Text and images can be well represented using deep neural networks. Jin et al. and Wang et al. [8, 12] applied it to fake news detection.

To overcome the limitation of learning shared representation of multi-modal data, Khattar et al. [13] proposed a Multi-modal Variational AutoEncoder. It is coupled with binary classifier features of text and image modalities with three components in the model, an encoder, decoder, and a fake news detection module. The model leverages state-of-the-art techniques with ~6% accuracy. Ajao et al. [14] used a hybrid of CNNs and LSTM-RNNs to identify fake news-related features without prior knowledge, achieving 82% accuracy. Jindal et al. [15] presented two novel datasets containing fake news text and image, using data augmentation to increase fake news data. Singhal et al. [16] perform fake news detection by introducing the SpotFake framework that exploits textual and visual features of news posts without considering subtasks such as event discriminator and modality correlations. The model increases the accuracy from previous approaches by 3.27% and 6.83% on Twitter and Sina Weibo datasets. TI-CNN has been proposed for fake news detection by Yang et al. [1] using Convolutional Networks on both textual and visual data. They have incorporated both explicit and latent features extracted for both the modalities using CNN layers.

A new challenge emerged to detect fake or computer-generated images with technological advancement in Generative Adversarial Networks. GANs pose a threat by allowing the creation of fake images and manipulations in existing images. Marra et al. [17] studied the performance of existing detectors that use conventional and deep learning methods, concluding higher efficiencies by deep learning detectors with 89% accuracy. They compared the performances of traditional and deep learning image forgery detectors on a dataset of 36302 images under compression and without compression, concluding that high accuracies are obtained on compressed data using deep networks like XceptionNet, InceptionV3, and DenseNet. In recent years, the yearly trend of published articles using deep networks for credibility analysis is represented in Figure 3. Figure 4 offers the percentage of fine-tuned CNN models in similar tasks.

By extracting event-invariant features, proposing event adversarial neural networks, Wang et al. [17] performed fake news detection on newly arrived events. Three tasks are completed, namely feature extraction, detection, and event discrimination. The study is conducted by ignoring features that are event-specific and considering just the shared features. It provides accuracies of 71.5% and 82.7% on Twitter and Weibo, respectively. Sabir et al. [18] detected image repurposing, i.e., manipulations in image meta-data on a self-proposed MEIR dataset that consists of real-world Flickr data. It proposes a multi-modal deep learning method that utilizes metadata and image information to identify modifications.

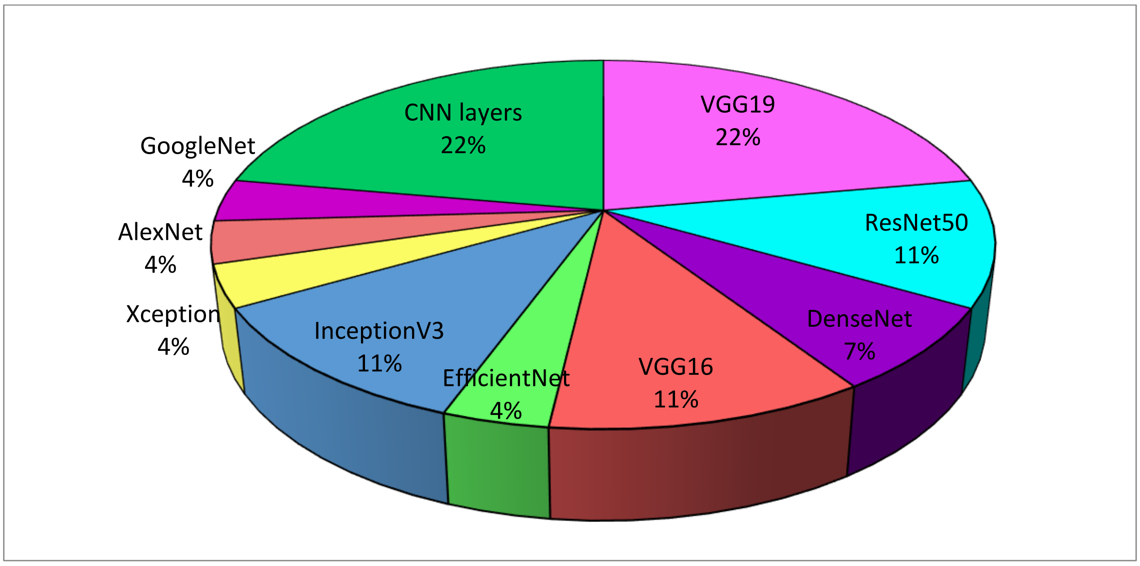

Figure 2: CNN architectures used in previous research

Pomari et al. [19] came up using CNNs and illumination maps in images to detect splicing in fake images with a colossal accuracy of more than 96%. Another approach used diverse modalities, including text, image, and source, to detect hoaxes [20]. Bayar and Stamm [21] developed a new convolutional layer that learns features from training data suppressing image features and highlighting manipulation features. This new approach can detect image manipulations with an accuracy of 99.10%. Lago et al. [22] performed the task using image forensics algorithms to see tampered images and a verification mechanism to check if the images are rightly mapped to textual news. In 2019, Cui et al. [23], a detection framework named SAME, exploits user comments and latent sentiments and uses an adversarial mechanism. Volkova et al. [24] performed a qualitative and quantitative analysis of fake news classification models, proposing a qualitative analysis tool ERRFILTER. Modalities analyzed are text, lexical and image inputs, and their combinations.

Table 1: ConvNet Architectures for credibility analysis of different data modalities

References	Modality	Task	Network	Model
[13]	Text, Image	Fake News Detection using MVAE	RNN, CNN	Bi-LSTM, VGG19
[14]	Text, Image	Fake News Detection on Twitter	Hybrid CNN, RNN	LSTM, CNN
[26]	Video	Face Manipulation Detection	RNN, CNN	ResNet50, DenseNet
[16]	Text, Image	Fake News Detection	RNN, CNN	BERT, VGG19
[1]	Text, Image	Fake News Detection	CNN	Bi-LSTM, CNN
[17]	Image	GAN-generated Fake Image Detection	CNN	DenseNet, InceptionV3, Xception
[12]	Text, Image	Fake News Detection	EANN	Text-CNN, VGG19
[18]	Image	Image Repurposing Detection	CNN	VGG19
[27]	Video	Fake Video Detection	RNN, CNN	LSTM, InceptionV3
[19]	Image	Image Splice Detection	CNN	ResNet50
[20]	Text, Image, Source	Hoax Detection	CNN	Deep CNN [1]
[21]	Image	Image Manipulation Detection	CNN	Proposed CNN
[22]	Text, Image	Image Trustworthiness Assessment in Online News	CNN	
[23]	Text, Image	Fake News Detection	CNN	LSTM, VGG16
[25]	Image	Classifying Computer-generated and Photographic Images	Modular CNN	VGG19
[8]	Text, Image, Video	Rumor Detection on Microblogs	Att-RNN	LSTM, VGG19
[29]	Image	Tampered Face Detection	Two Stream Neural Networks	GoogleNet, InceptionV3
[30]	Image	Classifying Computer Graphics and Natural Images	CNN	MLP

In image classification, Tariq et al. [25] detected fake face images generated by humans and machines using CNN-based models including VGG16, VGG19, ResNet, DenseNet, NASNet, XceptionNet, ShallowNet, and their ensembles. These neural networks detected

GANs and human-generated fake face images without using their metadata. The highest accuracies on various image sizes were obtained with Ensemble ShallowNet (V1& V3).

Sabir, Cheng, et al. [26] performed the detection in manipulations of faces in videos using recurrent convolutional models. These models have proved beneficial in utilizing temporal information in still images to detect tampered images improving the existing accuracies by up to 4.55%. Fake video detection has been performed by Guera and Delp [27], using a convolutional LSTM model on a large dataset of deep fake videos in which face swaps have been done. Papadopoulou et al. [28] verified real-time, user-generated online videos, YouTube videos taking their context into account. The information exploited includes video comments for textual data and metadata like video description, likes, dislikes, and uploader information.

CHAPTER 3

METHODOLOGY

This section elaborates on the architectures of the classification models utilized in this task. Proposed Coupled ConvNet is composed of Text-CNN module for textual fake news classification and Image-CNN module for visual fake news classification. We pre-process input data at their earlier stages in both modules and feed them to convolutional neural networks. This section explains the architectures and mathematical background of Text-CNN and other CNN models utilized in this work. Table 1 summarizes different Neural Network architectures used for the credibility analysis of data in various modalities.

3.1 TEXT-CNN

CNNs are widely used for visual tasks. For image classification, pixel information extracted from images is propagated as pixel values to consequent convolutional layers. Words are needed to be processed to make them understandable by a machine. A computing machine treats visual and textual data in the same manner as numeric data. The idea is to serve the machines with text in numeric data in the same way visible data is treated using pixel values. This task is performed by embedding words into vectors. Figure 5 details the various layers of text-CNN architecture.

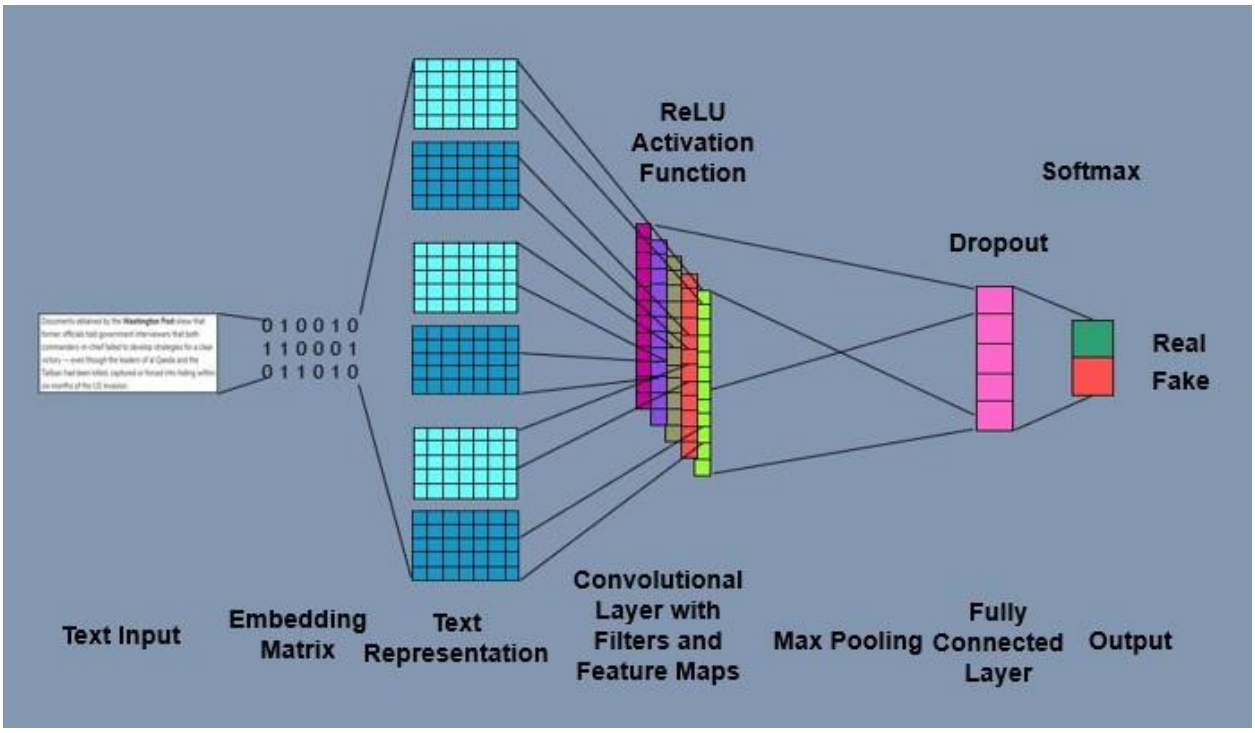
Documents obtained by the Washington Post show that former officials told government interviewers that both commanders-in-chief failed to develop strategies for a clear victory — even though the leaders of al Qaeda and the Taliban had been killed, captured or forced into hiding within six months of the US invasion.
0 1 0 0 1 0
1 1 0 0 0 1
0 1 1 0 1 0
ReLU Activation Function
Softmax
Dropout
Real
Fake
Text Input
Embedding Matrix
Text Representation
Convolutional Layer with Filters and Feature Maps
Max Pooling
Fully Connected Layer
Output

Figure 3: Text-CNN architecture

A fixed vector can thus represent each word in the sentence. These embedded vectors are then propagated through convolutional layers in the same way image data moves through the deep network. The consequent layers are of the same structure incorporating max-pooling, padding, activation function, fully connected layers, and dropout. It is mathematically represented in the form of a k-dimensional vector as $x_i \in R^k$, where x_i is the i^{th} word in a sentence.

Then, $x_{1:n} = x_1 \oplus x_2 \oplus \ldots \oplus x_n$, ($x_{1:n}$ is a sentence of length n) and $\oplus$ represents concatenation. The series of words $x_i, x_{i+1}, \ldots, x_{i+j}$ are concatenated as $x_{i:i+j}$. If h is the number of words, a filter w that is applied to the text generates a feature c_i from a word window $x_{i:i+h-1}$ where filter $w \in R^{hk}$and $c_i = f(w \cdot x_{i:i+h-1} + b)$ given b as a bias term and f, a non-linear function. The filter w is applied to every word window, producing a feature map $c = [c_1, c_2, \ldots, c_{n-h+1}]$ where $c \in k^{n-h+1}$. A max-pooling layer is applied next. It extracts the feature with maximum value in the feature map c which is expressed as $\hat{c} = \max\{c\}$. These features with maximum values are propagated further to fully connected layers passing them to a softmax layer for classification.

3.2 IMAGE-CNN

The fine-tuned CNN architectures provide good accuracies when it comes to extracting hidden image features and patterns. We implemented eight different CNN architectures AlexNet, Xception, VGG16, VGG19, ResNet50, MobileNetV2, InceptionV3, and DenseNet, for visual fake news detection. The designs of all fine-tuned image CNN models used in this work are described in the following section and represented in Figure 6.

AlexNet: AlexNet is a Convolutional Neural Network designed by Alex Krizhevsky in 2012 and won the ILSVRC challenge. The model displayed that depth in the network is necessary for efficient applications. Depth in the model contributed to providing high performance and became computationally costly, sufficed by using multiple GPUs. AlexNet architecture consists of 8 layers. The first five are convolutional layers, with each layer optionally being followed by a pooling layer and the last three layers are fully connected layers. The model prefers the ReLU activation function, owing to its advantage in training time over tanh or sigmoid functions. Overfitting encountered in AlexNet was reduced by data augmentation and using Dropouts that turn off neurons with a specified probability of 0.5.

VGG: Visual Geometry Group (VGG) won the ILSVRC 2014 competition. The group members, Karen Simonyan and Andrew Zisserman, experimenting with multiple numbers of layers in the deep network, released two versions of their model, VGG16 and VGG19, with 16 and 19 deep network layers each. They displayed that deeper networks with a more significant number of layers result in higher accuracy for image classification tasks. They replaced large kernel-sized filters of sizes 11×11 and 5×5 with smaller filters of size 3×3. Three fully-connected layers follow the convolutional layers following a softmax layer. ReLU is used as the non-linear activation function for hidden layers. The number of channels increases with a twice-factor from 64 in the first layer to 512 in the last layer. The increased depth makes VGG a network slower to train.

ResNet: Residual Neural Network is a network simplified by skipping layers introduced by Kaiming He in 2015. ResNet makes double and triple layers skips jumping across the network. This network makes training more comfortable and faster and reduces the vanishing gradient problem as there is a lesser number of layers in the network. It uses the ReLU activation function and Batch Normalization. Activations are reused from a previous layer until the current layer learns the weights.

Layers are indexed as $l-2$ to l for single skips in backward propagation and as l to $l+2$ for forward propagation. Given $k-1$ as the skip number, this can be generalized as $l-k$ for a backward skip and $l+k$ for a forward skip. A residual network building block with residual function $F(x)$ can be defined by the equations:

For equal dimensions of x and F, $$y = F(x, \{W_i\}) + x \tag{1}$$

and

For unequal dimensions, $$y = F(x, \{W_i\}) + W_s x \tag{2}$$

Here x is the input vector, and y is the output vector, $F(x, \{W_i\})$ is the residual mapping and W_s is a linear projection used for mapping dimensions.

Inception V3: The inception V3 model by Google for image classification was presented in ILSVRC 2015, providing a low error rate due to a 42-layer deep network. This model uses the factorization method to factorize a 5×5 convolution into two 3×3 convolutions. It reduces the parameters by 28%. Similarly, a set of one 1×3 and one 3×1 convolution can be replaced by a 3×3 convolution. The auxiliary loss tower in Inception V1 is used only on the last

17×17 layer as a regularizer in Inception V3. Inception V3 is observed to be much efficient than VGGNet in terms of computation cost.

Xception: Xception stands for "Extreme Inception," Its architecture is entirely based on depthwise separable convolutional layers. Its architecture consists of 36 convolutional layers (as 14 modules) followed by fully connected layers and a logistic regression layer. Except for the first and last modules, all convolutional layers have residual connections. The weight decay rate or L2 regularization of the Inception V3 model was improved to $1e - 5$, and the dropout layer used a probability of 0.5. The model does not incorporate the 'Auxiliary loss tower' that is optionally used in Inception V3 architecture.

DenseNet: DenseNets, introduced in 2018, are residual networks with various parallel skips. Each layer in a DenseNet is connected in a feed-forward manner to every other layer. The expression gives the total number of direct connections between the layers $\frac{L(L+1)}{2}$, where L is the number of layers. DenseNets do not require the learning of repeated feature maps and require a lesser number of parameters. They perform concatenation of feature maps instead of sum. Its equation can be stated as:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \tag{3}$$

Here x_l is the output of l^{th} layer and H_l is a non-linear transformation.

MobileNet V2: MobileNet V2, a type of CNN, was specially designed in 2019 for mobile devices based on inverted residual connections and bottleneck light-weight depthwise separable convolution layers. The first layer of MobileNet V2 is a convolutional layer with 32 filters. Nineteen residual bottleneck layers follow it. The kernel size used is 3×3, and the non-linear activation function used is ReLU6. The residual layers are used to make the model memory efficient. The bottleneck block operator used can be expressed as:

$$F(x) = \sum_{i=1}^{t} (A_i \circ N \circ B_i)(x) \tag{4}$$

where, A_i is a linear transformation, N is a non-linear transformation and B_i is a linear transformation to the output domain.

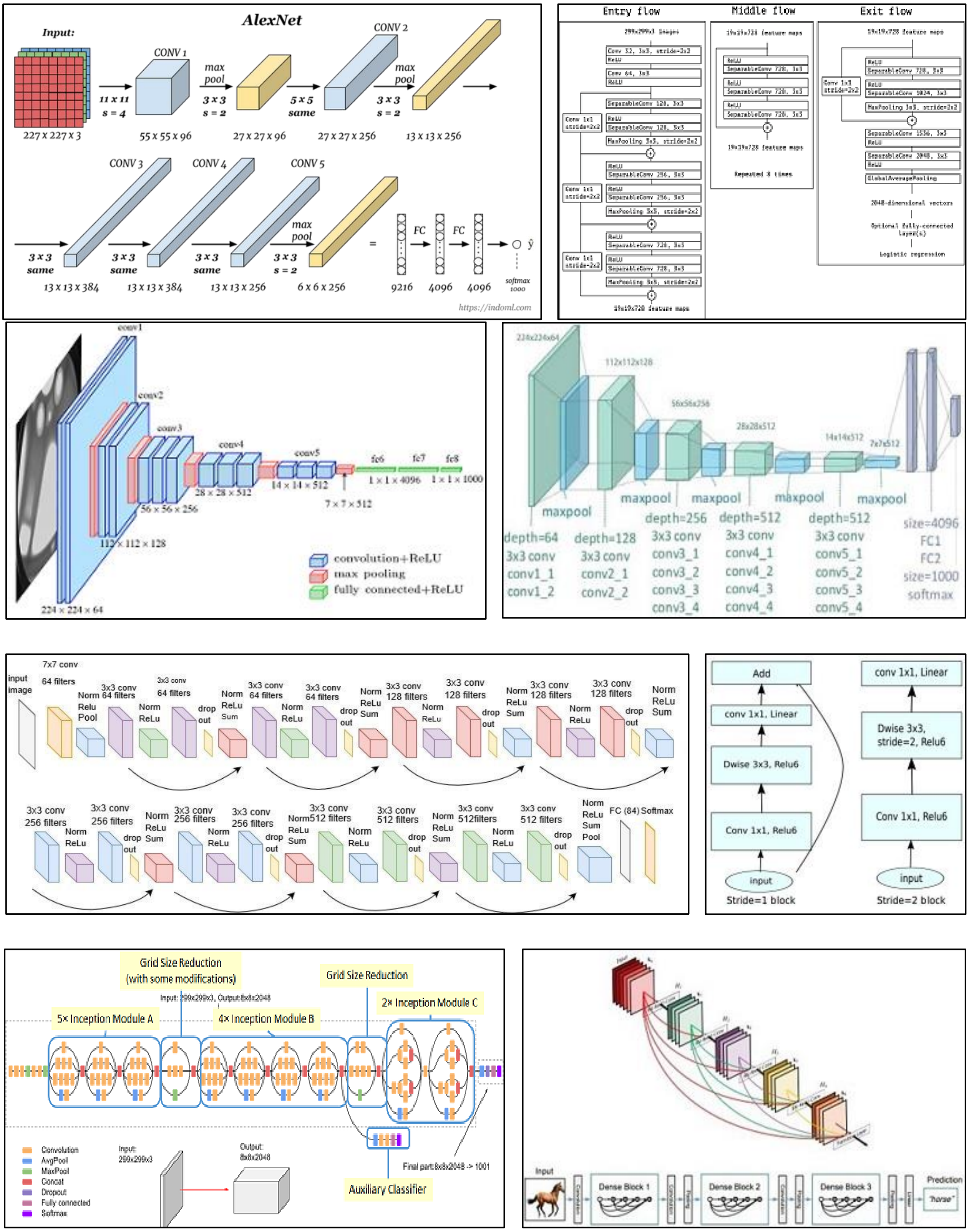

Figure 6: CNN model architectures: (a) AlexNet, (b) Xception, (c) VGG16, (d) VGG19, (e) ResNet50, (f) MobileNetV2, (g) InceptionV3, (h) DenseNet

3.3 PROPOSED COUPLED-CONVNET

The proposed approach to fake news detection extends the utilization of convolutional neural networks to a broader scale to automate fraudulent content detection on the web. Most of the existing literature is flooded with singular modality tasks where one of the present features are exploited. Most of the approaches are based on machine learning algorithms, while others use deep learning such as GRU, LSTM, Bi-LSTM, and other RNNs for text classification. We leverage this task by introducing a new text classification model using a convolutional neural network. With the onset of using visual features, pre-trained CNN networks are in wide use. The proposed image classification model is based on the usage of a pre-trained model with fine-tuning. Fake news detection tasks can be combined based on data modalities. Hence, the Coupled ConvNet introduced in this work is a hybrid two-stream convolutional architecture (based on text stream and image stream) is proposed, which are then combined using a late fusion technique. The architecture comprises of two streams (modules): Text Module (for textual classification) and Image Module (for visual classification). The architectures of these modules are explained in sections 4.1 and 4.2. The combination mechanism used in the proposed Coupled ConvNet is provided in section 4.3. The series of operations performed in both the modules is depicted in figure 7. Figure 8 represents the proposed Coupled ConvNet architecture.

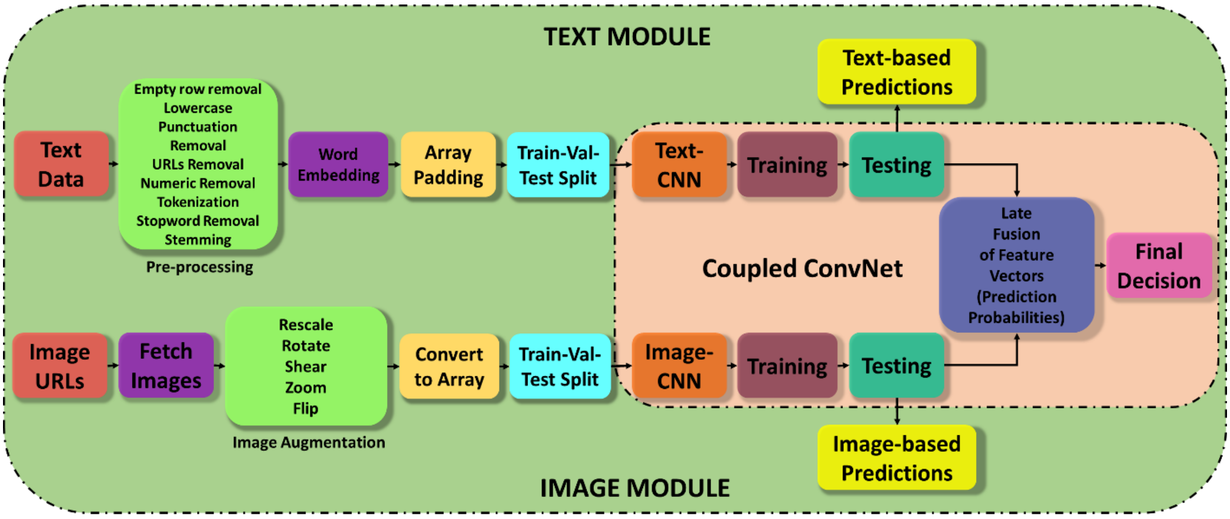

Figure 7: Sequence of operations performed

3.4 TEXT MODULE

A raw text dataset undergoes several refinements and analysis procedures before being affirmed for its realness. The first of those processes is pre-processing the text information. Then the word embeddings are generated for the textual content. Upon completion of this step, the embedded vectors are fed to a one-dimensional convolutional model. We then utilize the CNN model on textual data by applying convolutions on text vectors. A series of layers of convolution and pooling are generated to analyze the data features. Finally, all these layers are conjectured to provide a binary output of the data's information's authenticity. The results are obtained after training the data under multiple iterations of the proposed Text-CNN model.

We use only the 'title,' 'text,' and 'label' columns from the versatile information present in the datasets. Textual pre-processing involves the following steps: lowercase conversion, punctuation removal, URL removal, numeric value removal, data tokenization, stop-word removal, and stemming/lemmatization. In the next step, we perform Array Padding. Padding is done by calculating the maximum length from the most extended news item present in the array data. The text, which is shorter in length than the full content length, is padded with zeroes. The data is further split into the train, test, and validation sets. This processed data is now encoded, and text and title inputs are embedded using Glove embeddings. These embeddings are added next to the 1-D input layer. We then feed this data to the proposed CNN model. The proposed text classification model consists of three one-dimensional convolutional layers with ReLU activation function, each followed by a max-pooling layer. Subsequent layers are fully connected Dense and Dropout layers. After experimentation with different dropout values ranging between 0.2 to 0.8, the best results were portrayed by setting the value to 0.4 for both the dropout layers. A binary Sigmoid classifier is deployed to generate the predictions.

3.5 IMAGE MODULE

CNNs have shown considerable performance for various image classification tasks. They identify latent features without demanding any extra information. These latent features are present inside an image and are described as resolution, objects, pixel parameters, size of an image, etc. When the image data under examination is combined with other modalities such as text, it classifies real and fake news. For Image Analysis, the available image datasets are created as explained. The datasets consist of URLs of news pages. We use these URLs present in the database to scrape URLs of images present in those news pages, using BeautifulSoup. We download and then zip the fake and real photos from those newly obtained URLs into separate directories to our local access. These image URLs are also added to the datasets corresponding to their respective news. Data folders are uploaded to Google Drive, and the drive is mount to Google Colab. We use the split-folders module to divide the dataset into train, test, and validation sets with 80%, 10%, and 10% fake and real images, respectively, for TI-CNN and EMERGENT datasets. MICC-F220 dataset is split into 60%, 20%, and 20% for training, validation, and testing sets, respectively. A different proportion is used for MICC-F220. This difference in splitting ratios consists of 220 images, with 110 real and 110 fake images. Splitting this dataset into 8:1:1 leads to a minimal number of images in the validation and test sets. It creates a bias in the classification results. To avoid this bias and generate normalized results, this dataset has been split in a proportion that keeps a good number of images for validation and testing. After this, we perform Image Augmentation using ImageDataGenerator. Operations performed during augmentation include rescaling, rotation, shear, zoom, and flipping of images, which improves the quality of the datasets for usage. Image data is then fed to various mentioned CNN models for classification. The CNN training sequence is similar to that of the text convolution sequence except that in this case, two-dimensional convolutions are performed on visual (image) data. We feed visual data to various CNN models separately. The list of multiple models experimented with our data includes AlexNet, ResNet50, MobileNet, DenseNet, XceptionNet, InceptionV3, VGG16, and VGG19 [31-34]. Accuracy is determined after training the models for a specified number of epochs, and the result trends for training, test, and validation are observed.

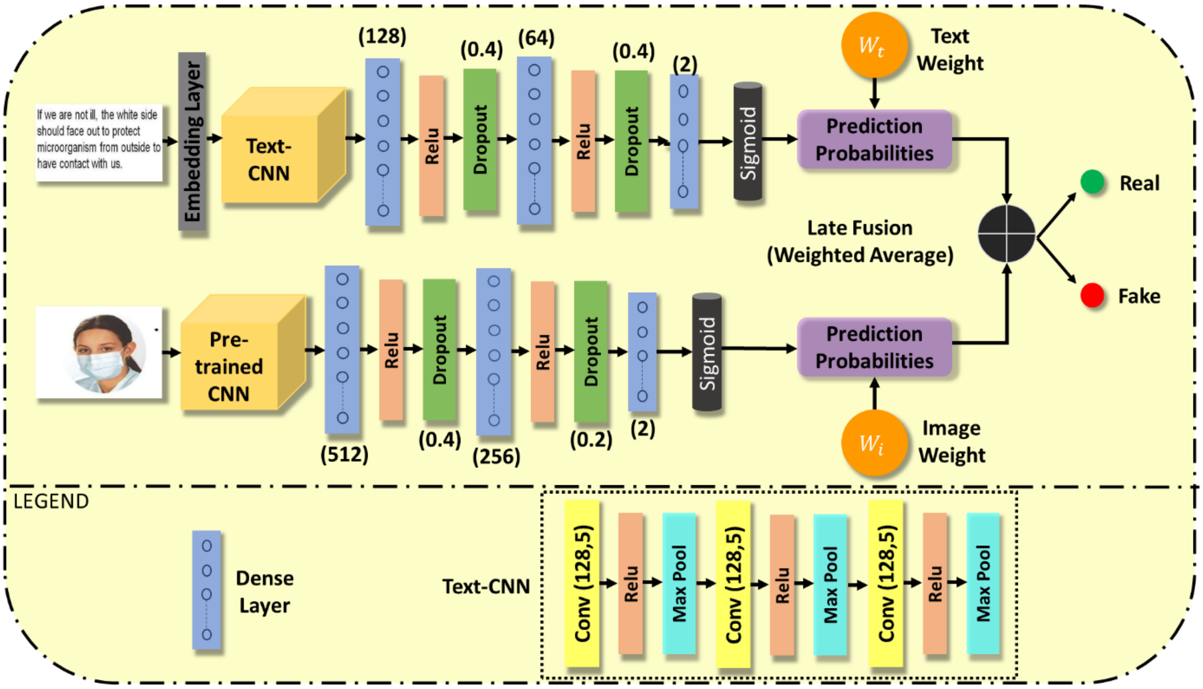

Figure 8: Proposed Coupled ConvNet Architecture

The proposed Image-CNN module uses one of the above-mentioned pre-trained models for each experiment. After adding a pre-trained model, a Dense layer of shape 512 with ReLU activation is added. Next, a Dropout layer of probability 0.4 is used. Another dense layer with shape 256 is used next. Subsequent layers incorporate a dropout layer of value 0.2 and a binary classification layer with a sigmoid activation function. The dense and dropout values are chosen decreasingly to avoid the immediate transition to the final classification layer. It allows the input to travel smoothly through the fully connected layers rather than directly jumping to the last layer. As observed during the experiments, using two dropout layers of value 0.4 and 0.2 reduces overfitting considerably and reduces the loss during the training phase, thereby increasing accuracy.

3.6 TEXT-IMAGE FUSION

Post-implementation of text and image classification modules separately, this segment fuses the outputs obtained. Prediction probabilities from both the modules are forwarded to a late fusion operation. Late fusion, a scalable and straightforward method, combines the features from multiple streams after the training phase. The decision vectors from each stream are combined using a suitable combinatorial operation. The proposed method uses a weighted fusion approach in which each modality is assigned a weight that determines the contribution of that modality in the final classification decision. Weights are chosen in a way such that maximum classification accuracy is obtained. For a fusion function $f : P_t, P_i \rightarrow P_c$ where $P_t \ and \ P_i$ are two different sets of prediction probabilities that denote the decisions of each stream, the combined probabilities indicated by P_c gives the output decisions after late fusion. P_c is calculated by adding the products of text and image prediction probabilities with their assigned weights W_t (text-weight) and W_i (image-weight). It is expressed as:

$$P_c = P_t * W_t + P_i * W_i \tag{5}$$

Choice of weights is made by experimenting with all possible combinations, varying the weight values between 0.1 to 0.9, with a difference of 0.1 unit. Text and image weights vary inversely. The variety of probabilities that produce the best result are used for each experiment. These weights have been described in table 2 in section 5.2.

CHAPTER 4

EXPERIMENTAL RESULT ANALYSIS

4.1 DATASETS AND PREPROCESSING

TI-CNN: With the availability of only a few good quality multi-modal datasets, we utilize the already collected dataset that is available online[1], used by Yang et al. [1] for a similar fake news classification task. This dataset contains 20,015 news items from websites, with 11,941 items being fake and 8,074 being real. The dataset is rich in terms of the wide range of details that it covers. We use all of these news items for the Text-CNN module using their title, text, and label information. For the Image-CNN module, we use image URLs obtained from the dataset in the 'main_img_url' column to scrape images from the web. The total number of images extracted from TICNN is 5733, constituting 2612 real news images and 3121 fake news images. The remaining URLs redirected to corrupted web pages or pages removed left us with an image dataset of a smaller size than their corresponding text items. TI-CNN dataset is used for experimentation in both Text-CNN and Image-CNN modules and later in the proposed Coupled ConvNet architecture.

EMERGENT: Another dataset experimented with is the EMERGENT (FNC) dataset created by Ferreira et al. [35], consisting of a total of 300 claims and 2595 associated articles. We polish this dataset by discarding duplicate news items and removing blank spaces. For the Image-CNN module, we use post URLs to extract image URLs and then scrape images from EMERGENT datasets' web-pages that led to a clean dataset of 1338 fake and 791 real images. We have made both of these image datasets publicly available. This dataset is also used in both the proposed individual modules and then in the proposed Coupled ConvNet architecture.

MICC-F220: Further, we used the MICC-F220 dataset by Amerini et al. [36] that consists of only real and tampered images, without any other form of data present. We use it with CNN models to identify whether an image is tampered with or original, in short, fake or real. Due to the lack of textual information, this dataset is solely employed in the proposed Image-CNN module. It is used to compare the efficiencies of utilized pre-trained CNN models within the proposed architecture.

[1] https://drive.google.com/file/d/0B3e3qZpPtccsMFo5bk9Ib3VCc2c/view

4.2 EXPERIMENTAL SETTINGS

All experiments have been performed on Google Colab that provides up to 13.53 GB of RAM. It also allocated us 12 GB NVIDIA Tesla K80 GPU hardware accelerator and python version 3. In Text-CNN, we employed RegexTokenizer to extract tokens from news titles and news texts. To reduce the words into their root forms, we used Porter Stemmer and WordNet Lemmatizer. We have utilized Glove representations for word embeddings used in Text-CNN. We have also applied one-dimensional convolutions on title and text and concatenated their layers. We used 0.4 and 0.8 as subsequent dropout values in the experiments. We used a batch size of 64 and have trained the model upon running for 250 epochs. For Image-CNN, we take the image input in size 224*224. Upon setting the dropout value to 0.2, the experiments exhibited a considerable increase in training accuracy. We have used Adam optimizer for all the given models. The batch size is set to 64 instances. The value of batch-size affects the training time of the model. The aim is to maximize the performance of classification models and minimize computation time. Choosing a batch-size less than 64 resulted in higher training time, which made the process slower. Whereas Google Colab did not accommodate a value greater than 64. Therefore, 64 is the perfect fit and is used as the batch size for both text and image modules. We have used binary cross-entropy loss for classifying the item into two categories: real and fake. In the combinatorial phase, weights for text and image features that provided the best classification accuracies were recorded and are as follows:

Table 2: Fusion Weights that provided maximum classification accuracies

Model	TI-CNN		EMERGENT	
	Text	Image	Text	Image
ResNet50	0.5	0.5	0.8	0.2
VGG16	0.5	0.5	0.5	0.5
VGG19	0.7	0.3	0.6	0.4
InceptionV3	0.8	0.2	0.7	0.3
DenseNet	0.5	0.5	0.8	0.2
Xception	0.5	0.5	0.5	0.5
AlexNet	0.6	0.4	0.7	0.3
MobileNet	0.5	0.5	0.5	0.5

4.3 RESULTS

This section presents the performance comparisons of all models used in our work for fake news classification on each of the three datasets. The scores are presented as accuracy, precision, recall, and F1-scores.

The comparison values of the Text-CNN module on two datasets, TI-CNN and EMERGENT indicate that CNNs exhibited an outstanding performance for classifying text-based fake news with 96.26% accuracy on TI-CNN and 93.56% accuracy on EMERGENT. Better scores were obtained on the TI-CNN dataset when compared to EMERGENT in all Text-CNN performance scores. It accounts for the larger size of TI-CNN data. More data aids in better training and hence produces better results. We portray performance comparison values for eight Image-CNN modules on TI-CNN and EMERGENT. VGG16 and VGG19 performed the best with 82.72% and 81.04% scores, respectively, on the TI-CNN dataset, followed by ResNet50 and MobileNet with 77.54% 73.37% accuracy, respectively. Other Image-CNN models scored below 63% accuracy on the TI-CNN dataset. The top four in terms of precision and F1 score were in the same order as the accuracy on the TI-CNN dataset with VGG16, VGG19, ResNet50, and MobileNet top four best performing models. In Recall scores, Inception V3 bagged 100%, followed by DenseNet, VGG16, and Xception, on the TI-CNN dataset. For the EMERGENT dataset, in terms of accuracy scores, ResNet50 and Xception secured 51.26% each (highest accuracy), followed by DenseNet and MobileNet with 48.65% 46.93%, respectively. VGG16 performed better on TI-CNN, whereas ResNet50 and Xception on the EMERGENT dataset indicate varying importance and reliance on different Image-CNN models regarding variations in the dataset. We show the performance of the eight Image-CNN models on the Image-only dataset MICC-F220. Xception with 100% accuracy, followed by VGG16 with 95.05% accuracy, VGG19 with 91.97%, and AlexNet with 91.54% accuracy, lead the table.

We provide the final output performance figures of the proposed Coupled ConvNet framework on the two datasets. Comparisons based on Accuracy, Precision, Recall, and F1 scores can be inferred from the table. To eliminate complexity in deciphering the best model or the most relevant text and Image multi-modal fake news detection, let us analyze the Accuracy score comparisons between the TI-CNN and EMERGENT datasets. The combination of Text-CNN with VGG16 performed the best on each of these datasets with

98.93% and 94.05% scores, respectively. While Text-CNN and VGG19 combination performed with 98.4% accuracy on TI-CNN as the second best, Text-CNN and MobileNet coupled ConvNet produced 93.98% accuracy on the EMERGENT dataset, being the second-best. Third and fourth-best performance on TI-CNN was observed with DenseNet and InceptionV3 with 97.86% and 97.65% accuracy, respectively, and on EMERGENT, ResNet50, and Xception produced 91.47% and 90.98% accuracy, respectively.

Weights produced the best classification results can be concluded to be 0.5 for both text and image. Text and image both offer an equal contribution to detecting fake news efficiently. In some cases, the participation can be discovered to be 7:3 for text and image data modalities. It highlights text being a necessary component for fake news detection. It is also evident that exploring visual modality is equally essential.

The MICC-F220 dataset consists of tampered and unaltered images. Images under the unaltered category have not been edited in any form, and thus it serves the purpose of efficiently distinguishing between real and fake images. We deduce that CNN models are highly accurate in detecting fake news where the text is classified based on their vector embeddings and images have been tampered with or edited. We propose using combinations of text and image CNN models to detect fake news using multiple textual and visual modalities. Hence, we provide performance comparisons of these models to make a witty selection for counterfeit news detection tasks. The accuracy obtained with the MICC-F220 dataset is as high as 100% using XceptionNet, and the lowest is 59.52% with the ResNet50 model. Other models have also demonstrated outstanding performance with high accuracy values. This performance highlights the need for larger visual and multi-modal datasets with distinguishable latent features.

It can be concluded that VGG16 is a consistent performer. Xception and MobileNet are observed to be the next best performers. Despite achieving 100% result with the MICC-F220 dataset, Xception displays average performance with the other two datasets. It can be regarded as being slightly inconsistent with datasets.

Table 3: Performance of Text-CNN Module on TI-CNN and EMERGENT

Dataset	Accuracy	Precision	Recall	F1-Score
TI-CNN	96.26	95.77	96.00	95.89
EMERGENT	93.56	94.07	89.35	93.12

Table 4: Performance of Image-CNN Module on TI-CNN and EMERGENT

Image Model	TI-CNN				EMERGENT			
	Accuracy	Precision	Recall	F1Score	Accuracy	Precision	Recall	F1Score
ResNet50	77.54	58.22	88.57	70.25	**51.26**	**58.59**	**76.82**	**66.26**
VGG16	82.72	63.49	97.65	77.26	45.18	51.29	58.56	54.77
VGG19	81.04	59.77	88.98	71.32	41.90	48.42	42.45	45.00
InceptionV3	58.76	09.18	100.00	16.81	43.54	50.28	54.41	52.89
DenseNet	60.00	11.40	97.96	20.43	48.65	52.93	63.71	57.25
Xception	62.57	10.51	97.62	18.98	**51.26**	57.19	68.42	62.59
AlexNet	59.44	48.32	91.69	59.87	43.62	50.71	48.64	49.71
MobileNet	73.37	55.66	79.46	65.46	46.93	55.18	52.53	53.48

Table 5: Performance of Image-CNN Module on MICC-F220

Image Model	MICC-F220			
	Accuracy	Precision	Recall	F1-Score
ResNet50	61.36	61.37	59.52	60.43
VGG16	95.05	95.15	78.26	85.88
VGG19	91.97	92.02	83.33	87.46
InceptionV3	91.01	91.01	90.63	90.82
DenseNet	89.63	89.61	92.00	90.79
Xception	100.00	100.00	93.75	96.78
AlexNet	91.54	91.52	95.00	93.22
MobileNet	82.82	82.73	100.0	90.55

Table 6: Performance of Coupled ConvNet Model on TI-CNN and EMERGENT

Text Model	Image Model	TI-CNN				EMERGENT			
		Accuracy	Precision	Recall	F1Score	Accuracy	Precision	Recall	F1Score
Text-CNN	ResNet50	96.90	96.48	96.71	96.59	91.47	91.90	88.95	87.96
	VGG16	98.93	98.21	99.22	98.71	94.05	90.08	86.72	86.12
	VGG19	98.40	97.18	98.96	98.06	89.12	89.11	85.80	85.49
	InceptionV3	97.65	97.65	97.19	97.42	89.88	89.63	86.44	85.88
	DenseNet	97.86	98.34	96.96	97.64	90.64	90.35	87.39	86.73
	Xception	97.22	94.36	98.92	96.59	90.98	91.77	88.02	87.97
	AlexNet	96.91	96.26	96.94	96.60	89.66	89.80	86.09	86.02
	MobileNet	97.54	97.41	97.18	97.29	93.98	91.48	86.22	86.55

Table 7: Accuracy Comparison of Image-CNN models on all datasets

Image Model	TI-CNN	EMERGENT	MICC-F220
ResNet50	77.54	51.26	61.36
VGG16	82.72	45.18	95.05
VGG19	81.04	41.90	91.97
InceptionV3	58.76	43.54	91.01
DenseNet	60.00	48.65	89.63
Xception	62.57	51.26	100.00
AlexNet	59.44	43.62	91.54
MobileNet	73.37	46.93	82.82

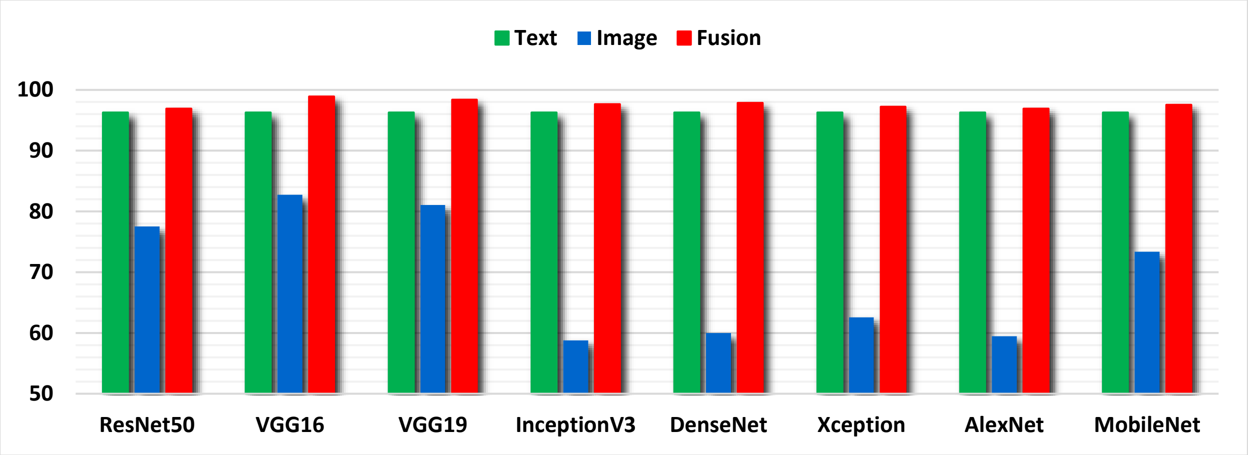

Figure 9: Accuracy Comparison on TI-CNN dataset

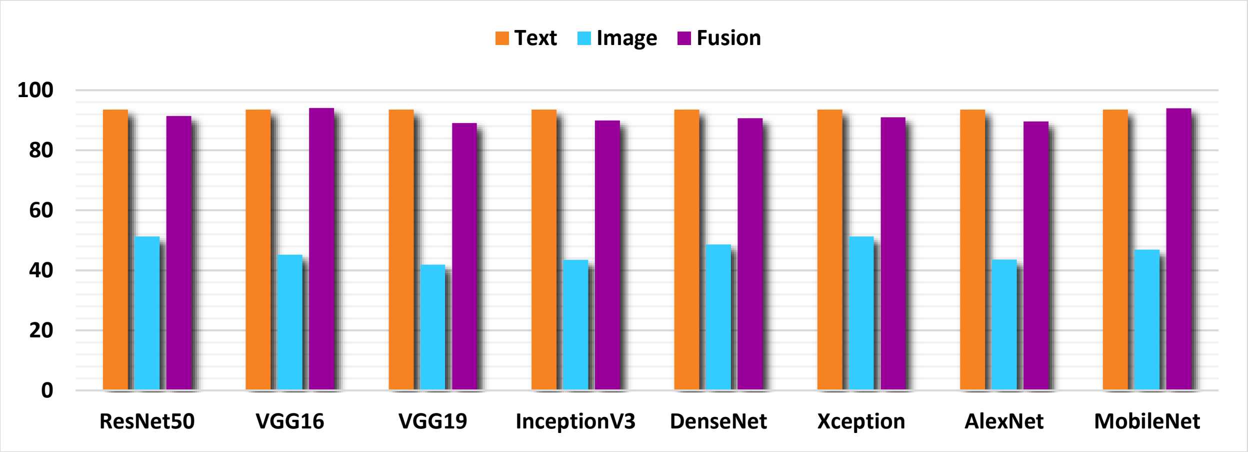

Figure 10: Accuracy Comparison on Emergent Dataset

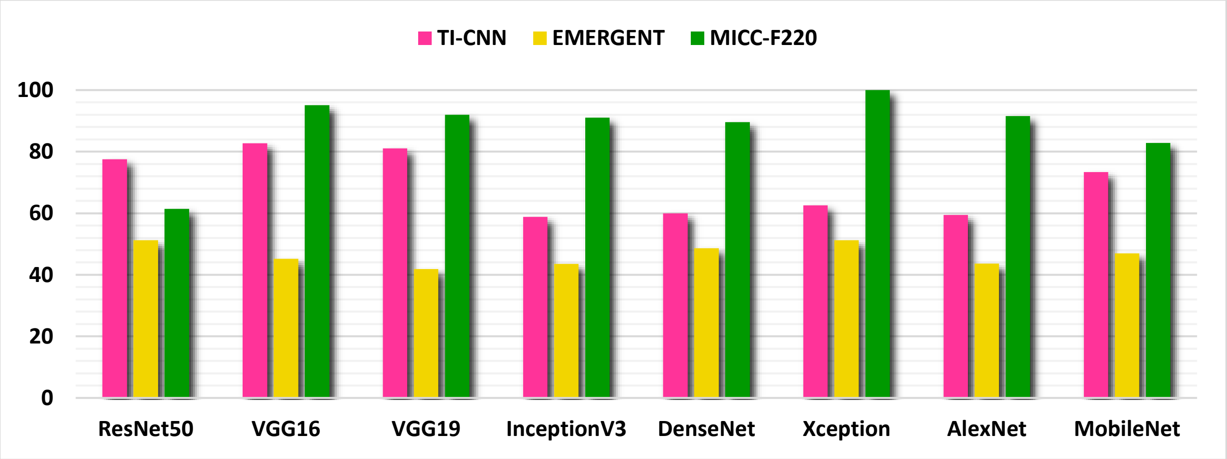

Figure 4: Accuracy Comparison of Image-CNNs on three datasets

We conclude that CNNs perform better when the dataset is comprised of all tampered images. Data with fake images where fake corresponds to false, tampered, old, misleading, and unrelated images perform somewhat lower as CNNs could detect only latent features. For utilizing features contained in all types of fake photos, multi-modal frameworks are needed which can incorporate elements contained in all kinds of counterfeit images. The above best performing models are likely to show better performance over larger training datasets.

4.4 BASELINE COMPARISON

We validate our results with both single modality textual and visual methods and multi-modal methods for a fair comparison of our proposed work with established baselines. We compare the results for each dataset separately. The proposed task being the first to examine Emergent on a visual basis, we establish a baseline for visual and multi-modal fake news detection on this dataset. Due to the absence of work performed in the visual domain, this task stands first, and hence, the comparison is provided for textual classification. The results for comparison are noted from those mentioned in the existing literature.

Table 8: Baseline comparison of TI-CNN dataset

Modality	Baseline	Method	P	R	F1
Textual	(Yang et al.) [1]	LR	57.03	41.14	47.80
		GRU	88.75	86.43	87.58
		LSTM	91.46	87.04	89.20
		Text-CNN	87.22	90.79	88.97
	Proposed Method	Text-CNN	**95.77**	**96.00**	**95.89**
Visual	(Yang et al.) [1]	CNN-image	53.87	42.15	47.29
	Proposed Method	ResNet50	58.22	88.57	70.25
		VGG16	63.49	97.65	**77.26**
		VGG19	**59.77**	88.98	71.32
		InceptionV3	09.18	**100.00**	16.81
		DenseNet	11.40	97.96	20.43
		Xception	10.51	97.62	18.98
		AlexNet	48.32	91.69	59.87
		MobileNet	55.66	79.46	65.46
Textual and Visual (Combined)	(Yang et al.) [1]	TI-CNN	92.20	92.77	92.10
	Proposed Method	ResNet50	96.48	96.71	96.59
		VGG16	**98.21**	**99.22**	**98.71**
		VGG19	97.18	98.96	98.06
		InceptionV3	97.65	97.19	97.42
		DenseNet	98.34	96.96	97.64
		Xception	94.36	98.92	96.59
		AlexNet	96.26	96.94	96.60
		MobileNet	97.41	97.18	97.29

TI-CNN: On this dataset, Yang et al. experimented with multiple text classification methods: Logistic Regression, GRU, LSTM, and Text-CNN [1]. For the visual domain, Yang et al. used image-CNN with a proposed architecture of convolutional layers. They created a TI-CNN dataset and performed text classification using the embedding layer and one-dimensional convolutional layer. Image convolution is achieved by using a model that contains three convolutional layers. Filter size is kept as 3×3. Thirty-two filters have been used, and the layers inculcate the ReLU activation function. All of our text and image models surpass the scores obtained by Yang et al. [1]. Individual text and image models proposed by us provide accuracies higher than those observed by Yang et al. In the multi-modal aspect, our approach obtains the highest F1-score of 98.71% using a combination of Text-CNN and VGG-16, which outperforms the state-of-the-art result by ~6%. It establishes the proposed work as a new baseline for multi-modal fake news detection.

Table 9: Baseline comparison on EMERGENT (FNC) dataset

Modality	Baseline	Method	Acc%
Textual	(Conforti et al.) [37]	Bi-LSTM	33.00
	(Bourgonje et al.) [38]	LR	89.59
	(Thorne et al.) [39]	Ensemble Method	90.89
	Our Method	Text-CNN	**93.56**

Table 10: Baseline comparison of MICC-F220 dataset

Modality	Baseline	Method	TPR%	FPR%
	(Uliyan et al.) [41]	Hessian Method	92.00	08.00
	(Uliyan et al.) [42]	Blur Detection	96.50	02.86
	(Doegar et al.) [40]	AlexNet	100.0	12.12
	(Amerini et al.) [36]	SIFT	100.0	08.00
	Our Method	ResNet50	59.52	**0.00**
		DenseNet	78.26	**0.00**
		AlexNet	83.33	**0.00**
		InceptionV3	90.63	83.33
		VGG16	92.00	**0.00**
		MobileNet	93.75	25.00
		VGG19	95.00	12.50
		Xception	**100.0**	**0.00**

EMERGENT: Experiments previously performed by researchers used FNC (FakeNewsChallenge) dataset, which has been derived from Emergent. We compare text-classification results of our model with the LSTM model used by Conforti et al. [37], Logistic Regression applied by Bourgonie et al. [38], and an ensemble of multiple methods deployed by Thorne et al. [39]. Usage of the Text-CNN classification model beats these established baselines, providing an accuracy of 93.56%. Visual fake news detection on this dataset has not been performed previously as the dataset was limited to textual information only. We leverage the task to a visual analysis by adding images extracted from page websites and provide a maximum of 51.26% accuracy using ResNet50 and Xception models.

MICC-F220: Earlier tasks on this dataset have incorporated image forgery detection techniques with Amerini et al. [36] demonstrating 100% TPR and 8% FPR. Most of our proposed model methods have displayed 0% False Positive Rate, and XceptionNet provides 100% True Positive Rate outperforming all other baselines. 0% FPR demonstrates that no fake samples were wrongly classified as real during the testing phase, and 100% TPR shows that all unaltered samples in the test set were classified into the correct class. A model that achieves 0% FPR and 100% TPR is a perfect classifier. With the proposed approach, the Xception model is the ideal classifier for this dataset, classifying all test samples into correct classes.

CHAPTER 5

CONCLUSION

A novel Coupled ConvNet architecture is proposed comprising of Text-CNN and Image-CNN modules. This work accomplishes fake news detection using several convolutional models on text and image data. Our first contribution provides image datasets for counterfeit news detection, which we have publicly available on Kaggle. We compare the performances of image classification models, namely AlexNet, ResNet50, DenseNet, MobileNet, Xception, InceptionV3, VGG-16, and VGG-19, on three real-world datasets TI-CNN, EMERGENT, and MICC-F220. Text-CNN module has been used over TI-CNN and EMERGENT and Image-CNN module on all of the above datasets. We have trained these models and obtained their training, validation, and testing accuracy scores. We utilized latent features for fake image classification and analyzed how well classification can be performed, comparing various efficiencies. All of our models have surpassed fake news detection baselines with high results. The proposed architecture provides a new fake news detection method using convolutional neural networks and establishes a new baseline in this domain. The source codes of the proposed work have been made publicly available. Our proposed model would function more efficiently on larger datasets. We intend to apply these models to larger datasets further. We are also motivated to tune further the parameters used in these models to enhance classification accuracy. Additionally, we focus on coming up with an efficient classification model based on CNN's with fine-tuned hyperparameters serving greater accuracies and better fake news detection.

RELATED PUBLICATIONS

[1] Raj, Chahat, and Meel Priyanka. "ConvNet frameworks for multi-modal fake news detection." Applied Intelligence (2021): 1-17.

REFERENCES

[1] Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2018). TI-CNN: Convolutional neural networks for fake news detection. arXiv preprint arXiv:1806.00749.

[2] Wu, K., Yang, S., & Zhu, K. Q. (2015, April). False rumors detection on sina weibo by propagation structures. In 2015 IEEE 31st international conference on data engineering (pp. 651-662). IEEE.

[3] Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K. F., & Cha, M. (2016). Detecting rumors from microblogs with recurrent neural networks.

[4] Ma, J., Gao, W., & Wong, K. F. (2019, May). Detect rumors on twitter by promoting information campaigns with generative adversarial learning. In The World Wide Web Conference (pp. 3049-3055).

[5] Jin, Z., Cao, J., Zhang, Y., Zhou, J., & Tian, Q. (2016). Novel visual and statistical image features for microblogs news verification. IEEE transactions on multimedia, 19(3), 598-608.

[6] Boididou, C., Andreadou, K., Papadopoulos, S., Dang-Nguyen, D. T., Boato, G., Riegler, M., & Kompatsiaris, Y. (2015). Verifying Multimedia Use at MediaEval 2015. MediaEval, 3(3), 7.

[7] C. Boididou, S. Papadopoulos, D.-T. Dang-Nguyen, G. Boato, M. Riegler, S. E. Middleton, A. Petlund, Y. Kompatsiaris et al., "Verifying multimedia use at mediaeval 2016." in MediaEval, 2016.

[8] Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. (2017, October). Multi-modal fusion with recurrent neural networks for rumor detection on microblogs. In Proceedings of the 25th ACM international conference on Multimedia (pp. 795-816).

[9] Qi, P., Cao, J., Yang, T., Guo, J., & Li, J. (2019, November). Exploiting multi-domain visual information for fake news detection. In 2019 IEEE International Conference on Data Mining (ICDM) (pp. 518-527). IEEE.

[10] Chen, T., Li, X., Yin, H., & Zhang, J. (2018, June). Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. In Pacific-Asia conference on knowledge discovery and data mining (pp. 40-52). Springer, Cham.

[11] Liu, Y., & Wu, Y. F. B. (2018, April). Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

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

[13] Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019, May). Mvae: Multi-modal variational autoencoder for fake news detection. In The World Wide Web Conference (pp. 2915-2921).

[14] Ajao, O., Bhowmik, D., & Zargari, S. (2018, July). Fake news identification on twitter with hybrid cnn and rnn models. In Proceedings of the 9th international conference on social media and society (pp. 226-230).

[15] Jindal, S., Sood, R., Singh, R., Vatsa, M., & Chakraborty, T. NewsBag: A Multi-modal Benchmark Dataset for Fake News Detection.

[16] Singhal, S., Shah, R. R., Chakraborty, T., Kumaraguru, P., & Satoh, S. I. (2019, September). SpotFake: A Multi-modal Framework for Fake News Detection. In 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM) (pp. 39-47). IEEE.

[17] Marra, F., Gragnaniello, D., Cozzolino, D., & Verdoliva, L. (2018, April). Detection of gan-generated fake images over social networks. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 384-389). IEEE.

[18] Sabir, E., AbdAlmageed, W., Wu, Y., & Natarajan, P. (2018, October). Deep multi-modal image-repurposing detection. In Proceedings of the 26th ACM international conference on Multimedia (pp. 1337-1345).

[19] Pomari, T., Ruppert, G., Rezende, E., Rocha, A., & Carvalho, T. (2018, October). Image splicing detection through illumination inconsistencies and deep learning. In 2018 25th IEEE International Conference on Image Processing (ICIP) (pp. 3788-3792). IEEE.

[20] Maigrot, C., Claveau, V., Kijak, E., & Sicre, R. (2016, October). Mediaeval 2016: A multi-modal system for verifying multimedia use task.

[21] Bayar, B., & Stamm, M. C. (2016, June). A deep learning approach to universal image manipulation detection using a new convolutional layer. In Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security (pp. 5-10).

[22] Lago, F., Phan, Q. T., & Boato, G. (2019). Visual and Textual Analysis for Image Trustworthiness Assessment within Online News. Security and Communication Networks, 2019.

[23] Cui, L., Wang, S., & Lee, D. (2019, August). SAME: sentiment-aware multi-modal embedding for detecting fake news. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 41-48).

[24] Volkova, S., Ayton, E., Arendt, D. L., Huang, Z., & Hutchinson, B. (2019, July). Explaining multi-modal deceptive news prediction models. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 13, pp. 659-662).

[25] Tariq, S., Lee, S., Kim, H., Shin, Y., & Woo, S. S. (2018, January). Detecting both machine and human created fake face images in the wild. In Proceedings of the 2nd international workshop on multimedia privacy and security (pp. 81-87).

[26] Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., & Natarajan, P. (2019). Recurrent convolutional strategies for face manipulation detection in videos. Interfaces (GUI), 3(1).

[27] Güera, D., & Delp, E. J. (2018, November). Deepfake video detection using recurrent neural networks. In 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-6). IEEE.

[28] Papadopoulou, O., Zampoglou, M., Papadopoulos, S., & Kompatsiaris, Y. (2017, June). Web video verification using contextual cues. In Proceedings of the 2nd International Workshop on Multimedia Forensics and Security (pp. 6-10).

[29] Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2017, July). Two-stream neural networks for tampered face detection. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 1831-1839). IEEE.

[30] Rahmouni, N., Nozick, V., Yamagishi, J., & Echizen, I. (2017, December). Distinguishing computer graphics from natural images using convolution neural networks. In 2017 IEEE Workshop on Information Forensics and Security (WIFS) (pp. 1-6). IEEE.

[31] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[32] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[33] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

[34] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[35] Ferreira, W., & Vlachos, A. (2016, June). Emergent: a novel dataset for stance classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 1163-1168).

[36] Amerini, I., Ballan, L., Caldelli, R., Del Bimbo, A., & Serra, G. (2011). A sift-based forensic method for copy–move attack detection and transformation recovery. *IEEE transactions on information forensics and* security, 6(3), 1099-1110.

[37] Conforti, C., Pilehvar, M. T., & Collier, N. (2018, November). Towards automatic fake news detection: cross-level stance detection in news articles. In Proceedings of the First Workshop on Fact Extraction and VERification (FEVER) (pp. 40-49).

[38] Bourgonje, P., Schneider, J. M., & Rehm, G. (2017, September). From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles. In Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism (pp. 84-89).

[39] Thorne, J., Chen, M., Myrianthous, G., Pu, J., Wang, X., & Vlachos, A. (2017, September). Fake news stance detection using stacked ensemble of classifiers. In Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism (pp. 80-83).

[40] Doegar, A., Dutta, M., & Gaurav, K. (2019). CNN based Image Forgery Detection using pre-trained AlexNet Model. International Journal of Computational Intelligence & IoT, 2(1).

[41] Uliyan, D. M., Jalab, H. A., & Wahab, A. W. A. (2015, August). Copy move image forgery detection using Hessian and center symmetric local binary pattern. In 2015 IEEE Conference on Open Systems (ICOS) (pp. 7-11). IEEE.

[42] Uliyan, D. M., Jalab, H. A., Wahab, A. W. A., Shivakumara, P., & Sadeghi, S. (2016). A novel forged blurred region detection system for image forensic applications. Expert Systems with Applications, 64, 1-10.