APPLYING DEEP LEARNING METHODS FOR SHORT TEXT ANALYSIS IN DISEASE CONTROL

A Thesis Presented to the Department of

Computer Science

African University of Science and Technology

In Partial Fulfilment of the Requirements for the Degree of

MASTER of Computer Science

By

Ezema Abraham Obinwanne

Abuja, Nigeria

December 2017.

CERTIFICATION

This is to certify that the thesis titled “Applying deep learning methods for short text analysis in disease control” submitted to the school of postgraduate studies, African University of Science and Technology (AUST) Abuja, Nigeria for the award of the Master's degree is a record of original research carried out by Ezema Abraham Obinwanne in the Department of Computer Science.

APPLYING DEEP LEARNING METHODS FOR SHORT TEXT ANALYSIS IN DISEASE CONTROL

By

Ezema Abraham Obinwanne

A THESIS APPROVED BY THE COMPUTER SCIENCE DEPARTMENT

RECOMMENDED: _____

Supervisor, Prof Ekpe Okorafor

_____

Co-supervisor

_____

Head, Department of Computer Science

APPROVED: _____

Chief Academic Officer

_____

December 9th, 2017

ABSTRACT

Developing countries have been plagued by recurrent cases of infectious disease outbreaks; coupled with the limitation of traditional disease control strategies, other approaches have been explored for disease control, with social media at the forefront. Data from this source is short, noisy, and informal in representation, thus, conventional natural language processing (NLP) methods are not well adapted for their structure. Hence, deep learning approaches for character-level word vector learning were explored to classify disease-related tweets, and an adaptive prediction model for outbreak monitoring was developed, using the Ebola virus disease as a case study. Our system showed better performance for the described task when compared with existing state-of-the-art architectures; also, our predictive model showed correlation with official reported cases, with early warning of fourteen days prior to official.

Keywords: Deep learning, NLP, disease control, short text analysis, word vector learning

ACKNOWLEDGEMENTS

To my little sister Mmesoma, who has been very upset because my absence from home affected her performance in assignments and projects, be patient, soon you will be so smart as to do all that hard work on your own.

My supervisor Prof. Ekpe Okorafor who has been most kind, helpful and understanding with my pace, I have learnt a lot from your methods and foresight. Everything I achieved in this work was possible because of the resources the management of the African University of Science and Technology, Abuja, made available. Nkoro Joseph Afamefula who was always available to brainstorm with me on my concepts, goals and limitations. I never met a colleague that selfless and creative. My friends Uche Fortune and Chiamaka Ikeme who helped me with data collection, I am grateful.

To my lovely parents and siblings who worry about my welfare even more than I do, your show of love makes me feel hopeful even in moments of despair; you all give my life meaning.

Dr. F.O Isiogugu and Chibugo, I could only wish there is more I can give you; and Dum Dum, alive or dead, you are still the best friend I could have.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

API	Application programming interface
CBOW	Continuous – bag–of–words
CDC	Centre for disease control
ConvNets	Convolutional neural networks
DeepLDC	Deep learning for disease control
DFE	Disease – free equilibrium
EPR	Epidemic preparedness response
EVD	Ebola virus disease
FC	Fully connected
IDSR	Integrated disease surveillance response
ILI	Influenza–like Illness
JSON	JavaScript object notation
NER	Named entity recognition
NoSQL	Not only structured Query Language
POS	Part of speech
NLP	Natural language processing
relu	Rectified linear unit
WHO	World health organization

CHAPTER 1
BACKGROUND TO THE STUDY

1.1 Introduction

The explosion of data in recent times has marked a new age for human society. Social media platforms such as Facebook, LinkedIn and Twitter offer a place for people to share information in real time; in 2016, Africa aggregated a total of 120 million Facebook users each month and the statistics for other social media platforms have shown similar growth (Fuseware & World Wide Worx, 2014; Parke, n.d.).

The increasing coverage of social media in Africa cannot be over-estimated; likewise, its potential in worthwhile projects of event monitoring, perception evaluation, information extraction and retrieval; thus, its recommendation in disease control strategies.

Notwithstanding the success of social media approaches in politics and business, and its prospects in epidemiology as being timely, collaborative and populace-centric; extensive analysis is imperative, as the tendency of information to be misconstrued is common when machines process natural language. Hence, the need for deep learning methods in social media analytics.

1.2 Background of the study

Africa has been plagued with epidemic disease outbreaks, preceding the 15th century; these occurrences tend to retard both the growth of the human population in the region and the development expectations (Spinage & House, 2012). Study of these cases shows a trend of recurrence of disease outbreaks in previously affected nations and a migration to neighbouring countries, which may be attributed to the ecological changes in the region (Kebede, Duales, Yokouide, & Alemu, 2010; Spinage & House, 2012).

The recurrence frequency and the associated mortality rate raises questions about the level of preparedness, surveillance efficacy and control efforts in place; thus, over the years, different approaches have been explored and fused with existing methods to mitigate disease propagation.

1.2.i Historical account of infectious disease outbreaks in Africa

Records show a number of viral diseases prevalent in Africa, such as cholera, meningitis, influenza, yellow fever, rickettsia, smallpox, HIV/AIDS, Lassa fever and Ebola. The total mortality score ranges in the millions, with Ebola and HIV/AIDS accounting for over 3 million recorded deaths (Spinage & House, 2012).

Table 1 shows a selected number of disease outbreak cases in Africa, with the estimated casualty scores.

Table 1: History of disease outbreaks in Africa

Diseases	year	First location in Africa	Casualty/cases
Small pox	100-B.c	Egypt	10% mortality rate
Influenza	1775 1891 1890	North Africa Coast of West Africa (Gambia, Ghana, Nigeria)	1,072 deaths 9,000 deaths 12,500 deaths
Cholera	1831 1865 1881	Egypt	100,000 deaths
Rickettsia	1893	Central Africa Republic Zimbabwe, Sierra Leone Ivory Coast	60%,45% & 7% Casualties respectively
Yellow fever	1900- 2013	West Africa	40,000 deaths
Ebola	1976	Congo(DRC)	1,200 deaths

1.2.ii Disease control

Walter R. Dowdle defined disease control as 'the reduction of disease incidence, prevalence, morbidity or mortality to a locally acceptable level as a result of deliberate efforts; continued interventions are required to maintain the reduction' (Dowdle, 1998).

Efforts in disease control interventions are targeted to reduce the contact rate of transmission, keep the infectious population low, shorten the infection span of the prevalent disease and obtain a disease-free equilibrium (DFE) in the population (Brauer & Castillo-Chavez, 2014).

Disease control involves:

- Prevention activities for disease event surveillance, preparedness, and rapid response.
- Eradication activities for isolation, treatment, and rehabilitation of infectious people.

A good disease control strategy involves both prevention and eradication, though they tend to overlap and can be carried out in varying orders during the intervention cycle.

A number of organizations in conjunction with the World Health Organization (WHO), Centre for Disease Control (CDC) and the health ministries of different countries are in affiliation for disease control purposes. These bodies have been active in epidemic preparedness and response (EPR) activities and integrated disease surveillance response (IDSR) strategies. Adapting traditional methods for data collection, disease identification, outbreak events and predictions, casualty estimation, and all the other metrics of disease outbreak.

1.2.ii.1 Social media in disease control

Due to the decision-pipeline involved in traditional methods, though unavoidable because of the sensitivity of health matters; information collection, validation and dissemination tend to be gradual. Judging from the fatality and transmission rate of the recent outbreaks of the Ebola virus disease

(EVD) and Lassa fever in 2014 and 2017, one can only conclude that a swift response would have averted the disaster and reduced the mortality rate.

Social media in disease control has been explored in recent years to get timely information for disease event surveillance, disease prevalence and detection in spatial locations, thus causing its potential in disease prediction and prevention to increase (Choi, Cho, Shim, & Woo, 2016).

Unlike documents which are over 1,000 characters, formal and follow the syntax of the target language, social media data is characterized by texts less than 500 characters and do not follow syntax hence, deep learning methods are favoured for processing their structure.

1.2.iii Deep learning in short text analysis

For short text analysis, word representation by word embedding has been adjudged most suitable for word similarity and sentence classification tasks (Komninos, 2016), which include: sentiment analysis, machine translation, question type classification, topic categorization; and word similarity for web queries and search processing.

Ground-breaking works by Bengio, Ducharme, Vincent, and Janvin (2003), Mikolov, Sutskever, Chen, Corrado, and Dean (2013), and Mikolov, Corrado, Chen, and Dean (2013) which applied multilayer convolutional neural networks to capture word semantics and syntactic properties paved the way for further advancement in the field of natural language processing (NLP).

In the event of disease outbreak, timely intervention is necessary to control the mortality scores; this can be achieved by effective disease monitoring and control. Information shared over social media and text messages could portend disease prevalence only if its curation is timely and accurate and the parameters estimated from the text data can be integrated into statistical models to forecast the disease dynamics. Though the possibilities with social media data are encouraging, it is also characterized by a high noise level and language processing barriers because of its informality.

Therefore, an efficient recommendation model for disease control must be robust in order to avoid false positive results, which can frustrate disease outbreak responses. To actualize this, word embedding is leveraged in short text analysis and a time series analysis of this data is used to analyse trends.

1.3 Aim and objectives of the study

1. Implement text analytics on short texts from social media based on deep learning techniques.

2. Achieve disease event surveillance by leveraging social media.

3. Develop a recommendation system for disease control decision-making.

1.4 Research scope

This work implements deep learning techniques in short text analysis for infectious disease surveillance and control.

Data related to the Ebola Virus Disease (EVD) over Twitter is mined and analysed for case clusters, underlying trends and variations upon which an adaptive disease control model can be designed.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

Text analytics is a field in natural language processing. It aims to extract the semantic, syntactic and contextual information of any written language (Farzindar & Inkpen, 2015). The desired contents are extracted from a large pool of data for the purpose of knowledge discovery. According to Vijayarani, Ilamathi and Nithya (2015), information extraction and retrieval are common processes peculiar to research areas of text mining, web mining, data mining, graph mining, multimedia mining and structural mining. In this chapter, the current methods in NLP will be treated together with their potential in disease control through social media resources. Also, to better describe the effect of this work and its relevance in existing natural sciences problem-interests, epidemiological compartments for infection transmission are discussed.

2.2 Text preprocessing: the rudiments of NLP

Written information comes as a continuous connection of letters to form words – words then form phrases and phrases then form sentences. These chunks of information are further identified as parts of speech and named entities. Before computer programs can make these distinctions, a number of processes are carried out on the raw text. They include the following:

Tokenization: The process used to get discrete words by breaking texts based on punctuation marks and white space occurrence. These words form the vocabulary content of the system.

Stop words elimination: These words are not the major language terms in documents, they usually comprise determiners and conjunctions. Stop words can be removed using a compiled list of words that add no extra purpose other than grammatical completeness to a document. Advanced approaches apply Zipf's law based on these criteria: words that appear only once in a document, words that appear least in the pool of documents (inverse document frequency), and words whose frequency of appearance are excessively high, according to Vijayarani et al. (2015).

Stemming: This connects the different nuances of a base word. These nuances could be in the form of plural forms, past tense or continuous tense.

Normalization: The different inflections words can take in different contexts is checked. The targets at this stage are hyphenated words, capitalization, acronyms, and in the case of query tasks, it takes care of spelling errors.

Part-of-Speech (POS) Tagging: Depending on a sentence, words tend to assume different functions, the aim of POS-tagging is to determine the part of speech that the words in each sentence take up. The task is semi-automated, stochastic and rule-based. Models are used to automatically tag the dataset and later it is checked for consistency by human annotators (Marcus, Santorini, & Marcinkiewicz, 1993). The Penn Treebank, containing a total of 4.5 million words, is the most common POS-tagger used. The accuracy of a tagger is judged not just by its annotation accuracy but also by its consistency, syntactic function, efficacy and redundancy rating in tags (Marcus et al., 1993).

Parsing: the task carried out in NLP parsing involves the use of a computer program to extract phrases and the functional dependencies in sentences using an annotated grammar set. This takes the form of phrase and clause analysis in linguistics using accepted annotated rules in the language.

Table 2.1: List of stop words in English language

STOP WORDS	
ARTICLES	the, a, an
QUANTIFIERS	all, few, many, each, any, both
DEMONSTRATIVES	this, that, these, those
POSSESSIVES	our, its, your
CONJUNCTIONS	for, but, so, or, if
PREPOSITIONS	with, to, on, of, in, by, at

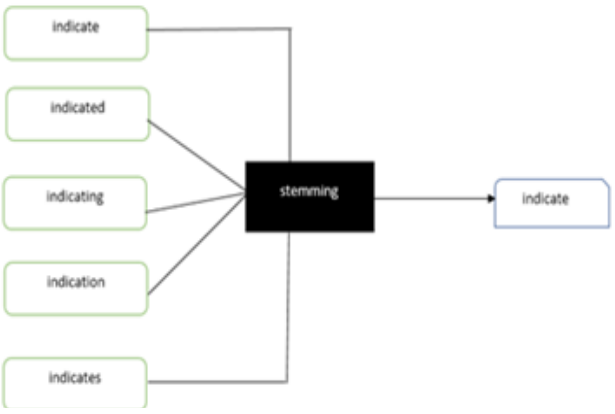

Figure 2.1: The stemming process

2.3 Text analysis for structured and unstructured data

Relevant data for a particular task comes in different forms, and knowledge of the kind of data to be analysed greatly affects the NLP steps to be adopted. Data collected for decision support, recommendation and ranking, prediction and forecast analysis, require a fine-grained representation of data in attributes and entities. When this data readily enters into a relational database, the process is easy and smooth, but with the target of these forms of analysis in recent times shifting to web pages, emails, product repositories and social media for movie ratings, health information, disaster events and trends, a shift from the norm has been observed, causing the data collection to be a 'cleaning' problem.

Recent research and industry goals have shifted towards technology that can manage unstructured data; the incidence of Bigdata and Not only Structured Query language (NoSQL) databases has been timely towards this endeavour. Various researchers (Baars & Kemper, 2008; Wakefield & Bean, 2005; Vivatrat & Falls, 2007) provide a structure for extracting data from an unstructured source into a structured representation, and represent this information in predetermined dimensions; this is termed information extraction. For advanced tasks in decision support systems, language parsing and word embedding are required to decipher and classify the knowledge embedded in unstructured data. For twitter data, the information returned is in JavaScript Object Notation (JSON) format, as a query selection of the fields that has to be performed to remove the unrequired data, e.g. noise, obtained from the twitter stream.

The result from extraction is data represented accordingly in a structured form in the database. Depending on the task required, sentiment analysis, data visualization, and further data analytics could be carried out on the data.

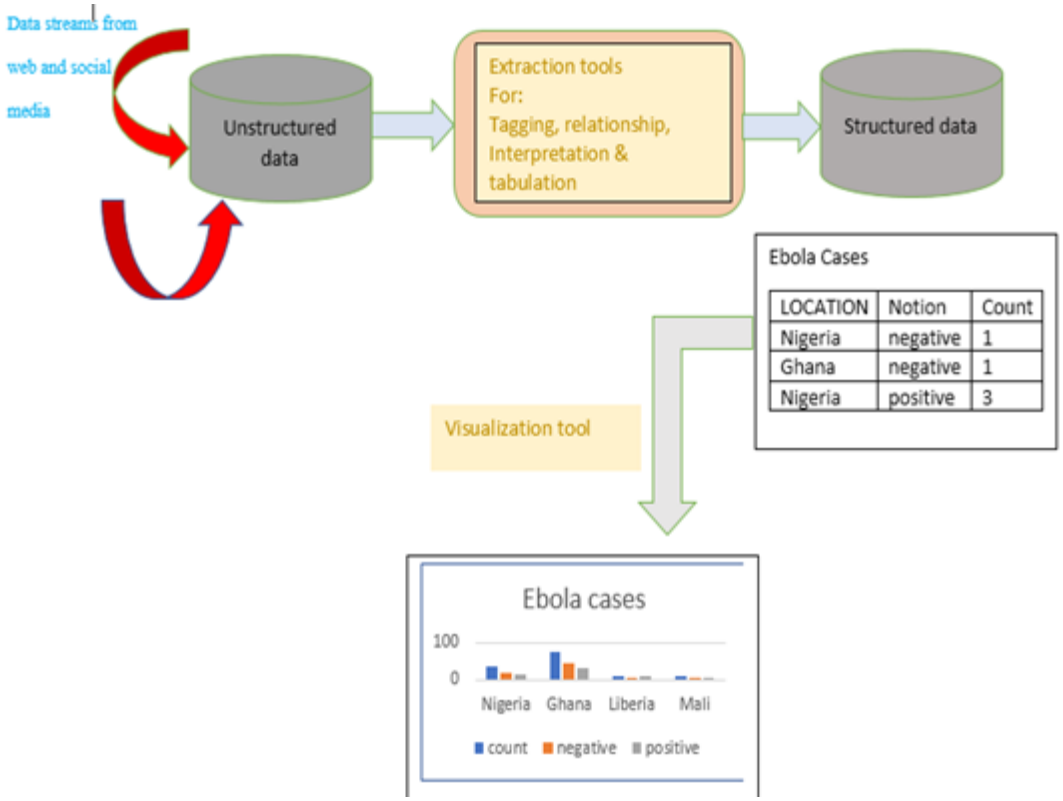

Figure 2.2: A framework for unstructured textual data analysis and conceptualization

2.4 Word vector learning

In order to process texts, it is necessary to represent words in a form that can be easily manipulated and interpreted by programs and machine learning algorithms. In word vector learning, words are mapped to real number values; in this form, programs in machines can easily carry out operations for information retrieval, information extraction, recommendation and ranking, document summarization, and other advanced NLP tasks.

Word vector learning is based on the distributional hypothesis; words which appear in the same context share similar meaning, as proposed by J.R Firth (1957): ‘words are known by the company they keep’. In line with this finding, words are represented in high-dimensional vectors such that the distance between them captures their similarities.

The output from text preprocessing is supplied as the input in this stage of NLP. This technique highlights language properties like: synonyms, antonyms, polysemy, named entity recognition (NER), and other word sense properties from vector manipulation.

2.4.i One-hot encoding

One way to initialize the values of word vectors is to assign a unique value to every word in the document. The length of the total unique words in the corpus is used as the row and column span of the word matrix. All the values of the matrix are assigned a value of zero, except the positions where the value of the row and column correspond, which are initialized with the value one; also known as sparse encoding.

- The limitations of the one-hot encoding include:
- It gives a sparse representation of word vectors and consumes a lot of space; when the corpus considered grows to large values, using this method for text processing becomes overwhelming and demanding.
- It does not capture word similarities; the distance evaluation gives the same value with every other word but itself.

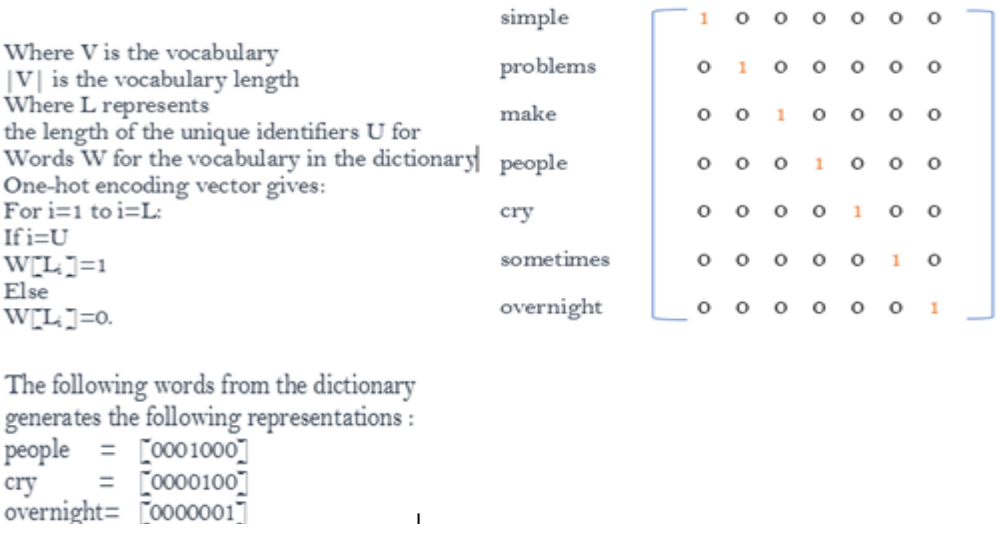

Figure 2.3: One-hot encoding steps with examples

It is used as the input of other text processing techniques which are better adapted to NLP and machine learning tasks.

2.4.ii Word embedding

Words are represented in continuous semantic similarity, based on the training corpus. The model captures the relationship between words in the corpus in relation to their appearance around some context words. This can be achieved in two ways, by means of count-based methods and by predictive models. The aforementioned approaches are known as dimension reduction methods in that their inputs are of large dimension but outputs are largely compressed to smaller dimensions following some algebraic methods and rules. Justifications for dimension reduction are:

- To manage the space that large dimension vectors cover, which most times are sparse.
- To reduce the complexity and computational cost of the model.
- Compressing the dimension brings out shared properties between words, which are not obvious in uncompressed vectors.
- By dimension reduction, a good representation can capture the similarities between the words in a language instead of having a bigger dimension to represent the millions of words that exist, as the output in these methods are dense matrixes.

2.4.ii.1 Count-based methods

In this method, the statistics of words co-occurring in a training corpus are taken and used to build the word vector representation of words. Given that words with similar meanings are used in the same context, the required features can be captured. A large co-occurrence matrix is first generated by comparing target terms in the vocabulary to the chosen context words from the training corpus. From practice, a number of considerations have to be made on the window size, window type, the distance measure, and the corpus size, as these parameters affect the word vector generated by different measures. Choosing the context words based on the high frequency words in the corpus proves to perform better than when based on the variance, truncated frequency and reliability; a

symmetric window type gives better performance; and a window size of 4 is a preferred choice (Levy & Bullinaria, 2000).

After the co-occurrence matrix is generated it is normalized by dividing the columns with the total number of times the target word appeared in the corpus to give a probability distribution of words. Using singular value decomposition, the dimension of the matrix is reduced, extracting the most important dimensions of the matrix; a dimension of 300 shows better performance as compared to higher and lower dimensions (Landauer & Dumais, 1997; Levy & Bullinaria, 2000). The Hellinger measures the distance between the word vector components to ascertain word similarity.

2.4.ii.2 Predictive models

Predictive models and counting methods share a number of term definitions; they also have similar real number concepts for word representations, probability distribution of words, target word determination in relation to context words, and window size and type. The major difference is that the word learning is done through the multilayers of a neural network by predicting the target word from context words, or context words from a target word in a sentence from the training corpus, applying the moving average concept. This involves more matrix operations and function transformations. The learning process is achieved by back propagation of the error in the predicted output for consecutive epochs, as compared to the expected output.

2.4.ii.3 Continuous bag-of-words model

Similar to the neural network architecture, the processes of forward propagation, activation of input and hidden layers, and training by backpropagation are used in the continuous bag of words (CBOW) model. The input to the model is a one-hot encoded vector of all the words in the vocabulary (V), as discussed earlier, and a dimension of N, which is also the dimension of the hidden layer. The objective of the CBOW model is to predict a target word in a sentence from a number of context words in the training set using a symmetric context type of four (Mikolov et al., 2013). Firstly, in forward propagation of the network, the context words that make up the inputs to the model are multiplied by the hidden layer weight matrix. The resulting hidden layer activation (h), is an average

of the corresponding rows of the vectors for all the context words (Bengio et al., 2003; Demeester, Rocktäschel, & Riedel, 2016; Mikolov et al., 2013; Pennington, Socher, & Manning, 2014; Rong, 2014); the hidden layer activation is defined by the equation below.

$$h=\frac{1}{C}\ W.(\sum_{i=1}^{C} X_i) \qquad (1)$$

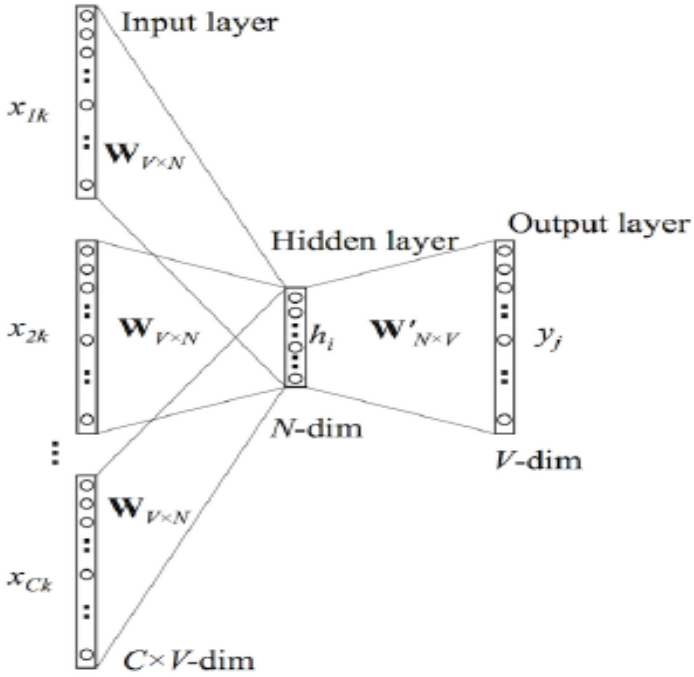

Figure 2.4: The continuous-bag-of-words architecture: multiple context words-to-1-target

Where:

C is the length of the list of context words

W is the weight matrix of the hidden layer of dimension [*V* x *N*]

X is the one-hot encoding of the context words

N is the number of neurons in the hidden layer

W' is the output weight matrix of dimension [*N* x *V*]

Y is the activation of the target word

The activation from the hidden layer is multiplied by the weights between the hidden layer and output layer; the activation function of the output layer is a softmax function. The aim of this step is to compute a score of all the words in the vocabulary in contrast to the target word (Rong, 2014). The error from the output is adjusted by stochastic gradient descent and propagated back in the layers to adjust the weights in the layers. The training process involves reducing the error to a global minimum; the gradient descent process points in the direction of this minimum using different measures to avoid leaping over the minimum and to converge faster (Andrychowicz, Denil, Gomez … de Freitas, 2016). The objective function used to predict the output word given other input words is negative log-likelihood (Chopra & Yadav, 2017; de Brébisson & Vincent, 2015; Geist, 2015).

The weights between the hidden layer and the output layer are the word representations after the learning process.

$$U_j = V'w_j{}^T . h \qquad (2)$$

Where:

$V'w_j$ is the column of the output weight matrix

h is the hidden layer activation calculated in (1)

U_j is the value of all the outputs

$$y_j = p(w_{yj} | w1, .. wc) = \frac{\exp(U_j)}{\sum_{j=1}^{V} \exp(U_j)} \qquad (3)$$

Where:

$p(w_{yj}|w1,..wc)$ is the softmax function

Y_j is the output of the network for all the words in the vocabulary

2.4.ii.4 Skip-gram model

The input of a skip-gram model is an input word of [1 x *V*] dimension and its output *C*, is made up of {*Wc*1, *Wc*2, . . ., *Wc*}, depending on the size of the set context words; this is a reverse of the CBOW architecture. It aims to generate the vector representation of context words from a target word, in order to capture word connections.

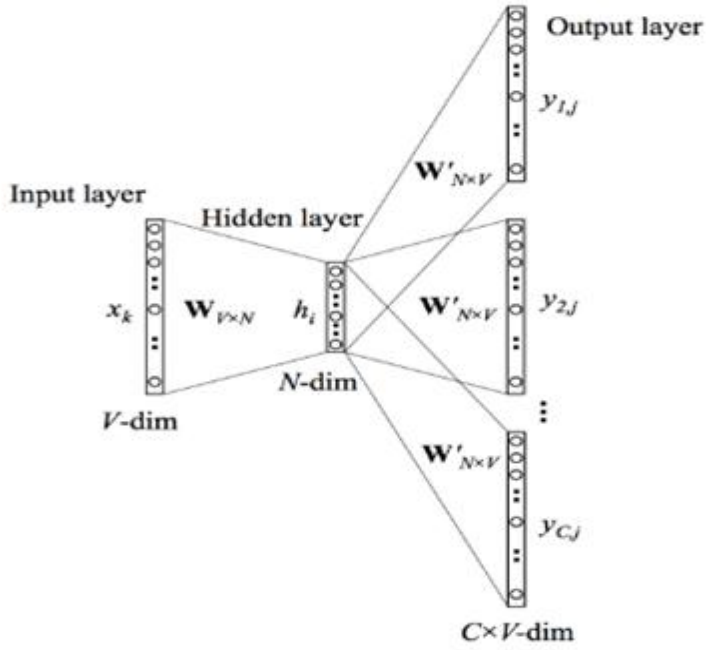

Figure 2.5: The skip-gram architecture: one target-to-multiple context words

Where:

C is the length of the list of context words

W is the weight matrix of the hidden layer of dimension [*V* x *N*]

X is the one-hot encoding of the context words

N is the number of neurons in the hidden layer

W' is the output weight matrix of dimension [*N* x *V*]

{*y1j…ycj*} are the output activations of the network

The weight matrix between the input layer and the hidden layer represents the context words predictions. During forward propagation, the hidden activations are replicated based on the number of context words set. The error in context words predictions is evaluated for each context word, the element-wise average of this error is used in backpropagation for weight updating across the layers and optimization is computed by stochastic gradient descent (Bengio et al., 2003b; Demeester, Rocktäschel, & Riedel, 2016; Komninos, 2016; Mikolov, Corrado, Chen, & Dean., 2013; Mikolov, Sutskever, Chen, Corrado, & Dean., 2013; Pennington, Socher, & Manning, 2014; Rong, 2014).

A number of variations to the prediction models exist, all in the view of better performance in word representation, memory requirement reduction and speed of training (Mikolov et al., 2013).

2.5 Disease compartment models in epidemiology

Disease compartment models in epidemiology are mathematical models that are paramount in the disease control sphere for epidemic and endemic diseases. They give an understanding of the causes of disease spread from which control strategies can be derived, based on empirical data (Van Den Driessche & Watmough, 2002; Varol, 2016). To apply these models, a number of assumptions have to be made in order to have a generalization of disease dynamics in a population, which in turn makes the system simpler and the parameters less complex to resolve (Brauer & Castillo-Chavez, 2012).

The assumptions for deterministic models described by state-space representation include:

- the mixing between the various compartments is homogenous,

- mass action incidence is the contact rate β considered: the rate at which people become infected by a disease at the instant just before the disease outbreak,
- the total number of people in the population is equal to *k*,
- where $N(_0)$ is the total number of people in the population before the disease outbreak.

Also, the compartment models show the threshold behaviour, which is determined by the basic reproductive number R_0. It is the average number of secondary infections caused by an average infected person (Brauer & Castillo-Chavez, 2012; Van Den Driessche & Watmough, 2002).

If $R_0 < 1$, the disease is eliminated from the population − the disease-free equilibrium (DFE)

$R_0 > 1$, the disease is an epidemic

$R_0 = 1$, the disease is endemic

It is used as a measure of the effect of control strategies, like vaccination and quarantine in the population.

2.5.i The SEIR model

This model is a perfect choice for the Ebola virus disease (EVD) analysis, which is considered in the scope of this work. EVD-infected people experience the exposed period that lasts for 2–21 days, known as the incubation period (Shen, Xiao, & Rong, 2015; Xia, Wang, Li … Jin, 2015), which is well captured in analysis (Brauer & Castillo-Chavez, 2012).

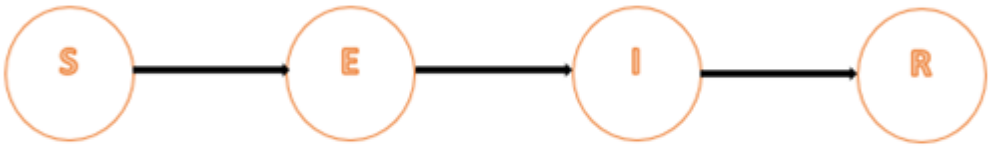

Figure 2.6: The SEIR flow representation

Where:

S is called the Susceptible class: members of the population that are disease free

E is the Exposed class: have been infected but cannot transmit infection yet

I is the Infectious class: manifest signs of infection and are capable of transmitting disease

R is the Removed class: people who have been removed due to death from the disease

$$S' = -\beta(N)S(I + (sE)E)$$

$$E' = \beta(N)S(I + (\varepsilon E)E) - \kappa I$$

$$I' = \kappa E - \alpha I$$

$$N' = -(1 - f)\alpha I. \quad (4)$$

$$R0 = \frac{K\beta(K)}{\alpha} + \varepsilon E \frac{K\beta(K)}{\kappa} \quad (5)$$

$$N = S + E + I + R \quad (6)$$

From (4) and (5), the parameter (εE) is the factor by which infectivity reduces during the exposed period because at that stage, even though one has been infected, EVD is not transmitted even when in contact with people; κ is the rate of transition from the exposed population to the Infected class; α is the rate of transition of the infected population to the reduced population; and *f* is the fraction of the infected population who recover after leaving the infectious class while (1-*f*) people die of the infection

2.5.ii SITR: The treatment model

Just like the SEIR model, the treatment model includes a treated class (*T*), obtainable when there is treatment available for people with the infectious disease and the motive is to ascertain how effective the treatment is (Brauer & Castillo-Chavez, 2012).

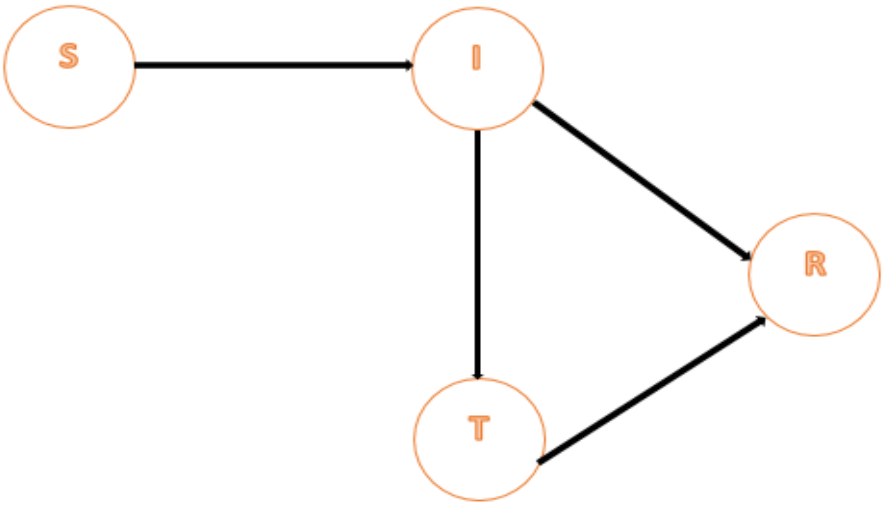

Figure 2.7: The SITR flow representation

$$S' = -\beta(N)S[I + \partial T]$$

$$I' = \beta(N)S[I + \partial T] - (\alpha + \Upsilon)I$$

$$T' = \Upsilon I - \eta T \quad (7)$$

$$R_0 = \frac{\beta K}{\alpha+\Upsilon} + \frac{\Upsilon}{\alpha+\Upsilon}\frac{\partial \beta K}{\eta}. \quad (8)$$

$$N' = -(1-f)\alpha I - (1-f_T)\eta T. \quad (9)$$

From (7) and (8), at a fraction *Y*, infectives are selected for treatment; the treatment in turn reduces infectivity by a fraction ∂; and the rate of transition from the treated compartment is η.

2.6 Related work

In an effort to prevent an epidemic, early disease detection is necessary in infectious disease control; this can be carried out using a number of disease surveillance mechanisms (Chan, Brewer, Madoff … Brownstein, 2010; Hashimoto, Murakami, Taniguchi, & Nagai, 2000; Simonsen, Gog, Olson, &

Viboud, 2016). Recent work has explored different novel mechanisms (Bansal, Chowell, Simonsen, Vespignani, & Viboud, 2016; Chan et al., 2010; Simonsen et al., 2016; Troppy, Haney, Cocoros, Cranston, & DeMaria, 2014); a few have been related to social media adaptations. Compared to other methods, social media can be adopted in surveillance schemes that promise early disease detection in the populace, efficient and affordable disease monitoring and good spatial analysis of disease spread (Bansal et al., 2016; Choi, Cho, Shim, & Woo, 2016; Simonsen et al., 2016).

2.6. i Leveraging Social Media and Emotion Classification

Ofoghi, Mann, and Verspoor (2016) proposed an approach using emotion classification of health-related incidents for disease outbreaks and monitoring. According to the work, signs of an outbreak can be noticed by detecting a shift in the nature of discussions about a specific disease; exploiting the relationship of public mood and socio-economic events. A shift in discussions entails a change in the distribution of emotions around supposedly affected areas immediately after and prior to the incidents. Using this approach, the need for text classification into predefined categories of discussion is avoided.

2.6. ii Establishing correlation with CDC reports and spurious data effects for influenza

This article by Culotta (2010), took the approach of correlating retrieved data from Twitter, together with the Centre for Disease Control and prevention (CDC) reports. After retrieval, spurious data was filtered using regression analysis. Then the proposed approach was tested on influenza-like illnesses (ILI).

ILI data was curated over Twitter, the results showed that spurious messages do not affect performance and there exists a strong correlation between Twitter data on ILI and CDC reports. The conclusion drawn was that the cases of misleading tweets are not significant to lead to false alarms for ILI.

2.6.iii Time series modelling and the temporal diversity of different infectious diseases

Kanhabua and Nejdl (2013) achieved time series modelling and the temporal diversity of different infectious diseases by matching tweets that had relevant keywords and location for 14 different diseases. Tweets had to have a location mentioned in the post body, have geolocation or the location was provided in the user's profile. Outbreak-related topics were identified by unsupervised clustering and irrelevant posts were filtered out, based on a repository. Similarities of documents in the generated time series clusters were performed using the Jaccard coefficient and the result differed for all the diseases at different locations; it also showed high correlation for mumps, Ebola, botulism and Enterohemorrhagic Escherichia coli (EHEC).

2.6. iv Integrating computational epidemiology models and social media

Zhao, Chen, Chen, Wang, Lu, and Ramakrishnan (2016) discussed epidemiology models that analyse the underlying mechanisms in disease transmission but lack information about the exact contact network and mixing within the population. Social media data is used to fill the gap and the strengths of these two approaches are explored in this work. The shortcomings of social media sources for establishing accurate statistics of infected people in the public was highlighted and its advantage for spatiotemporal information exploited. Tweets were used as inputs to the compartment model and in forecasting trends.

2.6. v A hybrid approach involving traditional and big data in disease surveillance

Simonsen et al. (2016) gives an overview of disease surveillance, the traditional methods and the gap still to be filled, like the lag-in reports and updates. The potential of big data applications to improve methods for timely, flexible and local tracking of diseases was detailed. Although digital methods seem to be flexible, systems based solely on them tend to fail, citing Google Flu Trend; therefore, a fusion of traditional methods and the big data approach makes for a robust and better adapted approach to disease surveillance and was advocated.

CHAPTER 3
ANALYSIS AND PROPOSED METHODS

3.1 Introduction

Good disease control entails efficient monitoring of media sources, though social media shows encouraging possibilities with real time information sharing and wide coverage, this information source can only be a viable option if the structure of the data processed from it can be well represented in analysis.

3.2 Data collection

Data is the major requirement to consider when designing a model. For generic applications, many samples of relevant annotated corpus are readily available; unlike for task-specific objectives, data has to be sourced, annotated and aggregated from the original stages.

Twitter was selected as the target social media platform for this work because of its wide coverage, provision for data streaming and search requests. Different events in different communities elicit different reactions from the populace; these variations and trends in the tweets generated in the locations of interest are extracted and captured to design a disease monitoring system.

3.2.i Obtaining historic data of disease outbreak

The dataset used is based on EVD-related publicly accessible tweets from Liberia, Sierra Leone, Nigeria and the United States of America, which had past EVD outbreaks; by using the Twitter advanced search provision accessed through the URL https://twitter.com/search-advanced, data from search results between March 2014 and March 2016 was collated. Some issues of location conflicts in the location field were resolved by using the geo-code of specific locations in the countries of interest; this enhanced the search radius for results.

3.2.ii Data labelling

Tweets were collated under the following classes:

1. Outbreak-related tweets: Contained tweets about the disease spread in a particular area, deaths, new cases confirmed or suspected, reports of illness, flu or fever, new outbreaks and description of diseases and its effects.

2. Prevention, control and management related tweets: Contained tweets about control efforts, possible and rumoured cures, end of viral spread, and prevention techniques.

3. Not directly related tweets: Tweets about Ebola workshops, seminars, research papers and study, and Ebola jokes. Some tweets that showed class overlapping were disjoined and classified into appropriate classes to aid subsequent classification tasks.

3.3 Text analysis

The approach used for text analysis stems from the approach in Zhang and LeCun (2015); character-level text analysis defeats the taxing requirement of a large vocabulary collection, word preprocessing, rule-based grammar definitions and extensive feature engineering for natural language processing (NLP) tasks. This text analysis type is more adapted to social platforms, which are grossly informal in language syntax, with a conversation style that involves the use of vernacular, contain links, and punctuation to express perception on matters.

3.3.i Text preprocessing

A predefined dimension of the 26 letters in English language, all the numbers, punctuation marks and newline, which sums up to 70 characters, was defined as the character field as shown below:

a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, !, “,”, #, $, %, &, (,), *, +, ‘,’, -, ., /, :, ;, <, =, >, ?, @, [, \\,], ^, _, `, {, |, }, ~, /n,\.

This captures all the characters represented in the target language and Twitter conversations.

3.3.ii One-hot encoding for character embedding

This is similar to one-hot encoding for word embedding described in Section 2.3.1, the only difference being the use of characters instead of words. For short text analysis purposes that comprise a limit of about 500 characters, the dimension of the one-hot vector has to be explicitly defined. Twitter limits the text field of tweets to about 140 characters; therefore, it is expedient in this work to consider this length limitation when implementing the input field quantization.

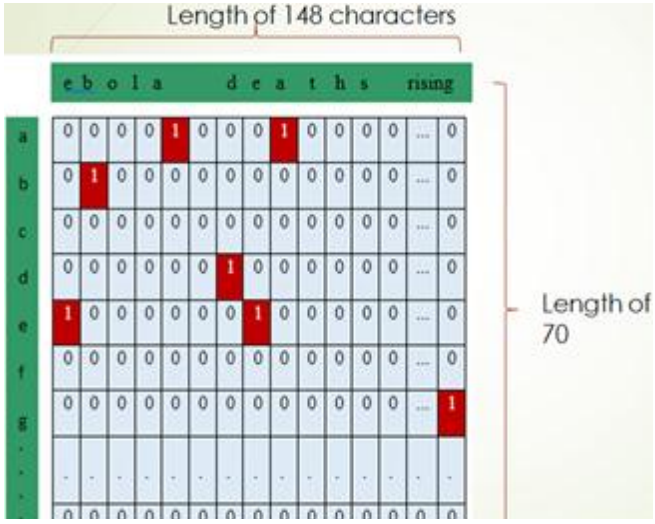

Figure 3.1: One-hot Vector for character embedding

Shown in Figure 3.1, the vertical plane indicates the predefined character fields that are listed in Section 3.2.1; the vertical field is the Ebola-related tweets, and the cells of value '1' in red correspond to the matching tweet input characters and the order in which they appear. This forms the input sequence to the text processing model.

3.3.iii Deep learning for text classification

Deep learning models have shown the ability to learn hierarchical features, and they also show outstanding performance in data intensive tasks for computer vision, speech recognition and natural language processing. In this work, we apply deep learning methods to analyze Twitter chatter on EVD; with the size of the available data, which does not add up to millions of data samples as used

in conventional deep learning implementations, we envisage comparable results that can be applied in disease prediction analysis.

3.3.iii.1 Convolutional neural networks (ConvNets)

ConvNets uses neural network architecture, as explained in Section 2.3.2.2, together with the following added concepts from (Krizhevsky, Sutskever, & Hinton, 2012):

Convolutional layer: This comprises a receptive field of weights that connects to a specific area of the input matrix; this area of reception is propagated along subsequent layers of the network. In practice, the weights capture features by dot multiplication with the affected input area during forward propagation and these weight values are adjusted in the course of training, based on calculated training error.

Kernel: Convolution is achieved by the settings of the kernel, which iterates over the entire input dimension. The value of the kernel size and stride determine how the input matrix is iterated.

Pooling layer: By means of a sliding window determined by the desired values of size and stride, the pooling layer performs down sampling over an input field. In max pooling, the highest value in the affected field is retained in the transformed vector, while other values are discarded; in average sampling, an average of the values is estimated and retained. Pooling is done to reduce the total number of computations carried out, by choosing the most important features to be learnt.

Padding: When padding is set, extra fields of zeros are appended to the input width; this is a technique used to achieve an unchanged dimension after a convolutional layer.

The concepts described above result in the transformation of the input matrix and are carried out at various layers of a ConvNet. The following formula describes these transformations and can be used to keep track of the variations in the dimension of the input field per layer (Zhang & LeCun, 2015):

Given an input sentence $i(x) \in [1, l] \rightarrow R$ and a discrete kernel function $f(x) \in [1, k] \rightarrow R$, the convolution $h(y) \in \left[1, \left\lfloor \frac{l-k}{s} \right\rfloor + 1\right] \rightarrow R$ between $f(x)$ and $i(x)$ is given as:

$$h(y) = \sum_{x=1}^{k} f(x).\, i(y.s - x + c). \qquad (10)$$

where *s* is the number of strides and $c = k - s + 1$ is an offset constant.

3.4 Model specifications

The proposed model for short text analysis stems from the Crepe model used in Zhang and LeCun (2015); character-level text analysis was used to show state-of-the-art performance in text categorization for documents in English and Chinese.

3.4.i Existing approach

Learning features of text from the characters contained in sentences unlike the word-level that tokenize words and generates a dictionary of all possible words. The length of the embedding used was predefined at 1014 characters. Features were learnt by a 9-layer ConvNet.

Architecture:

[conv-relu-pooling] *2 → [conv-relu] *3 → [conv-relu-pooling] → [FC-relu] *2 → Output scores.

FC-stands for fully connected layers, max pooling for pooling operation.

Specifications:

LAYER	Large features	Small features	Kernel	Pool
1	1024	256	7	3
2	1024	256	7	3
3	1024	256	3	N/A
4	1024	256	3	N/A
5	1024	256	3	N/A
6	1024	256	3	3

Layer	Output units large	Output units small
S7	2048	1024
8	2048	1024
9	Depends on the problem	Depends on the problem

3.4.ii Proposed approaches

Using a predefined dimension of 1014 characters on the Twitter dataset would result in an excess of 867 null-valued fields in the input matrix, at best case, thereby, using limited system resources on computations that are out of range. A model better adapted to Twitter text length is proposed for text analysis – topic categorization is the scope of this work.

3.4.ii.1 Model architecture

i. Clipped our input length to 148 characters

ii. [conv–relu] *2 →[pooling]*3 → [FC–relu] *2→[Fully connected Layer–softmax] → score of 3

iii. This amounts to 5 layers.

3.4.ii.2 Model specifications

Table 3.2: Model specifications

LAYER	NO OF FILTERS	KERNEL LENGTH	POOL
1	**256**	7	**N/A**
2	**256**	7	**2,2,2**

LAYER	OUTPUT UNITS
3	**1024**
4	**1024**
5	**3**

Applying the definition (10) from 3.2.3.1, the input transformation across the layers of the ConvNet can be calculated as follows:

Layer 1:

$$142 = \frac{148 - 7 + 2(0)}{1} + 1$$

Layer 2:

$$136 = \frac{142 - 7 + 2(0)}{1} + 1$$

Pooling 1:

$$68 = \frac{136 - 2 + 2(0)}{2} + 1$$

Pooling 2:

$$34 = \frac{68 - 2 + 2(0)}{2} + 1$$

Pooling 3:

$$17 = \frac{34 - 2 + 2(0)}{2} + 1$$

(17 x 256) parameters are fed into the fully connected layers.

3.4.iii Model structure

Figure 3.3: Dimension transformation across the network

3.4.iii.1 Disease outbreak prediction

To achieve a disease prediction model for an early warning system, we generated a system that can detect an abnormal increase in outbreak-related tweets in form of spikes from the Twitter stream; which serves as an event-trigger for our system. It is worth noting that an increase in outbreak event chatter does not necessarily translate to an outbreak; therefore, a related spike in the prevention-, control-and management-related tweets is also evaluated to make decisions on possible outbreak points using time series analysis concepts.

Our system uses the concept of an additive sliding window across previous counts (14-day range considered) of streamed tweets to measure an aggregated signal level against a new entry. This sliding window comparison to a new stream takes the form of a measure of the momentum of tweets. This measure serves the function of monitoring the rate of the rise in the level of new tweet counts, as compared to the old momentum, forming two momentum graphs. At any point where the new count momentum exceeds the aggregated momentum, an outbreak is possible; this point is called a cross-boundary.

We carried out further tests at this point to confirm an outbreak, as a form of cross-validation, by comparing the momentum of prevention- and control-related tweets. An opposite behaviour: where the count level is less than the aggregate level confirms a new outbreak, according to our system. The decision in the confirmatory test is based on the observed trend in collected data that in event of a new outbreak, less data about the other two classes is generated and in cases where there is an even or relative count across all classes, conversations tend to be on general views, lack first-hand outbreak content, and do not indicate panic.

The Twitter streaming API gives search results for the past seven days; this stream of data is used periodically for disease event surveillance. Also, for a social media approach to be timely and preemptive, the monitoring period has to be short enough to capture events of outbreak cases. The method we propose is adaptive and checks multicriteria for recommendation.

Algorithm Pseudocode

//This algorithm implements our disease prediction model, first detects spikes or cross boundaries in data streams which it uses to confirm cases of new disease outbreaks where T_0=outbreak related streams & T_p=prevention related streams

```
T_stream = (0: T_o ;1: T_p; 2: T_nr)
Check= 0
T_o= [0,0,0]
T_p= [0,0,0]

while T_stream:
        T_o [] =window (T_o [], T_o)
        T_p [] =window (T_p [], T_p)
        if (check=0)
                S_1=SlowStartSum (T_o [])
        Else
                S_1=Sum (T_o [])

        Event_1 = Compare (S, T_o)

        if (Event_1 = 0)
                if (check=0)
                        S_2=SlowStartSum (T_p [])
                Else
                        S_2=Sum (T_p [])

                Event_2 = Compare (S_2, T_p)
                if (Event_2=1)
                        print ("Detected case of new possible outbreak at check #", check)
                 else
                        print ("Test did not confirm at check #", check)
```

Subroutines Implementation

//window implements the sliding window, sum the total of the window elements and compare implements the check for a spike or cross boundary

```
window (W [], w)
        W. append (w)
        del W [0]
        return W []

Sum (S [])
        i_sum← 0
        for (i=0, i<len (S [])
                i_sum=+S [i]
        return i_sum

SlowStartSum (S [])
        i_sum← 1
        for (i=0, i<len (S [])
                i_sum=+S [i]
        return i_sum

Compare (S, T)
        if (S – T >= 0)
                result←1
        else
                result←0
        return result
```

Based on the motivation of this work and the initiative behind its introduction, we call this model Deep Learning for Disease Control (DeepLDC).

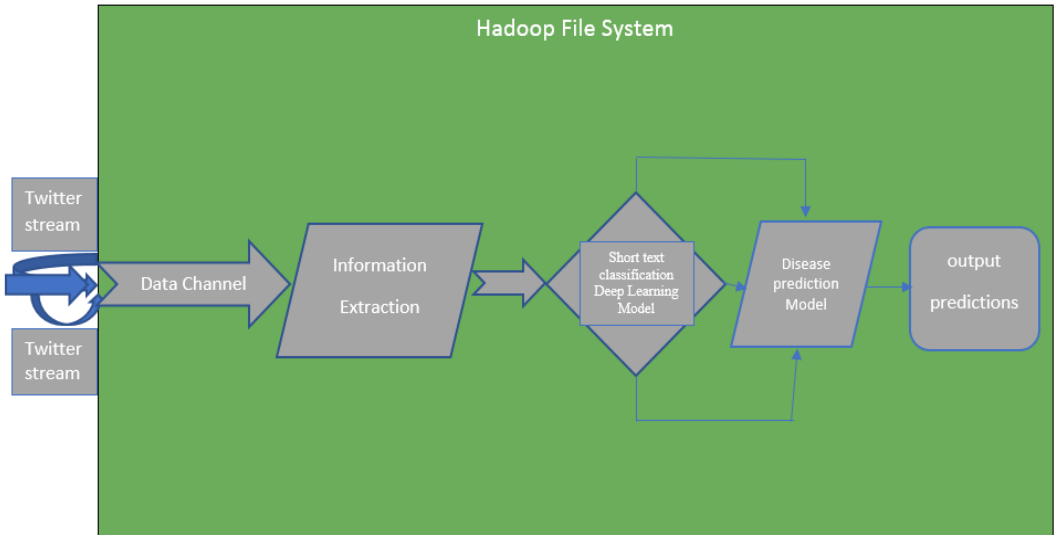

Figure 3.8: System architecture

CHAPTER 4
IMPLEMENTATION AND SIMULATION

4.1 Introduction

The Convolutional Neural Network (ConvNet) model, discussed in Section 3.3.2.2, used for short text analysis was implemented and its performance was measured by benchmark evaluations. All the results presented in this chapter were based on the data collated from Twitter using the procedures outlined in Section 3.1.1 and the output classes are as defined in Section 3.1.2.

4.2 Implementation procedure

Some key factors were set in defining the of the ConvNet module; we used non-overlapping max pooling throughout the pooling layers; rectified linear unit (relu), a non-saturating linearity function was implemented after each convolution layer which gives the network more freedom to adjust weights during training (Glorot, Bordes, & Bengio, 2011). We randomly initialized weights from a truncated normal distribution defined by mean=0.00 and standard deviation=0.05, and the network was trained using stochastic gradient descent. A minibatch of 80 was used for the input text sequence. Adaptive delta (adadelta) was implemented for optimization with learning rate initialized as 1, with a reduction rate of 0.6 after two non-improving training loss errors. A dropout of 0.5 was inserted between the fully connected layers to regularize and check overfitting of the network (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014; Tobergte & Curtis, 2013).

4.3 Deployment and use case

The training data set comprised the following:

- outbreak related data (labelled #0): 350 tweets;
- prevention, control, and management data (labelled #1): 350 tweets;
- not directly related (labelled #2): 350 tweets.

The validation/test data set:

- outbreak-related data (labelled #0): 91 randomly selected tweets from the dataset;
- prevention control and management data (labelled #1): 90 randomly selected tweets from the dataset;
- not directly related (labelled #2): 89 randomly selected tweets from the dataset.

Dataset imbalance for the three categories was avoided, as some categories gave way more search results than others; conscious effort was made to scale them to equal number.

4.4 Evaluation of results

Table 4.1: DeepLDC performance evaluation

Class	Precision	Recall	f1-score	Support
0	0.56	0.63	0.59	91
1	0.65	0.54	0.59	90
2	0.52	0.55	0.54	89
Average/total	0.58	0.57	0.57	
Validation accuracy:	63.16%			

In Figure 4.1, vertical axis represents true classes, top to bottom, while the horizontal axis represents predicted values.

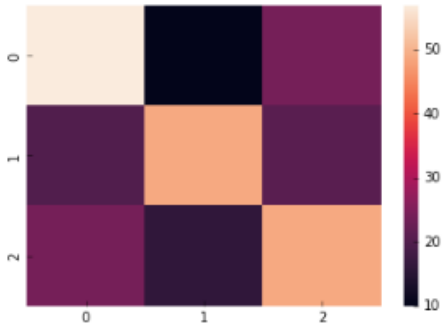

Figure 4.1: Confusion matrix

Class 0: From the confusion matrix shown in Figure 4.1, out of 91 randomly chosen test cases, predicted about 58 tweets correctly; this was estimated using the colour scale and axis of the map (colour code on top left against label '0' horizontal axis).

Class 1: out of 90, predicted about 50 correctly (colour code at 2nd row, 2nd column against label '1' on horizontal axis).

Class 2: out of 89, predicted about 50 correctly (3rd row, 3rd column against label '2' on horizontal axis).

4.4.i Correlation with reported cases

Our model picked out the following possible disease outbreak points leveraging the algorithm in 3.2.2.4:

1. Data point 22, which corresponds to 23–29 July 2014.

2. Data point 35, which corresponds to 21–27 October 2014 tweets gathered on an outbreak in Mali.

3. Other data points did not meet the condition of our confirmatory test as their prevention-related level does not lag the prevention window at those points.

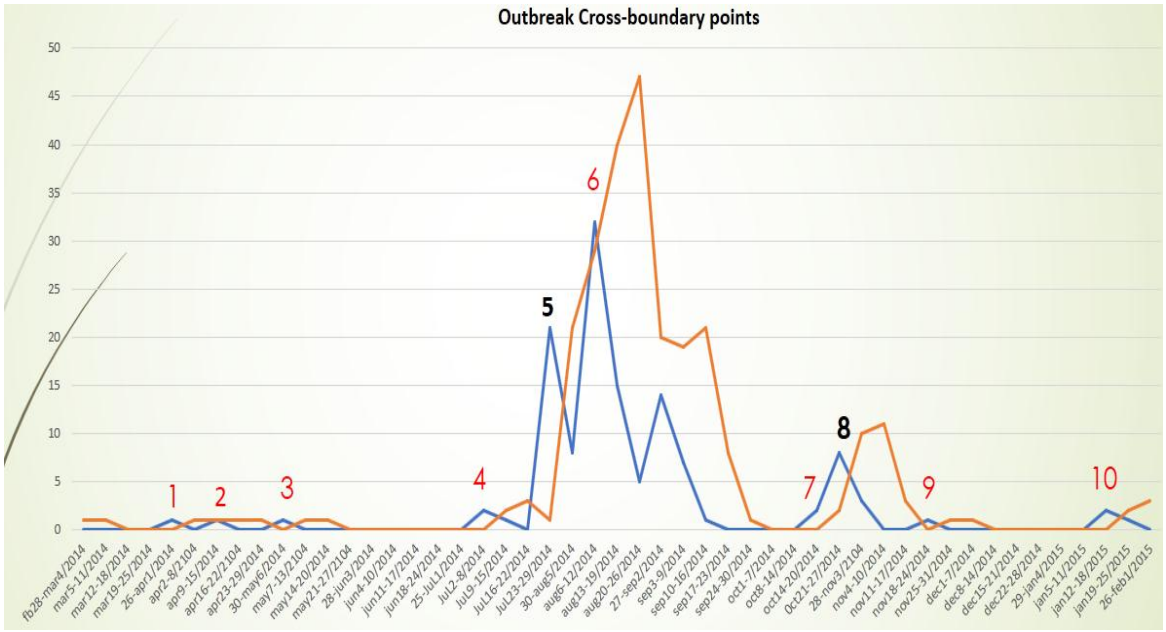

Figure 4.2: Disease prediction graph using a sliding window size of 2 for Nigeria (feb2014-feb2015)

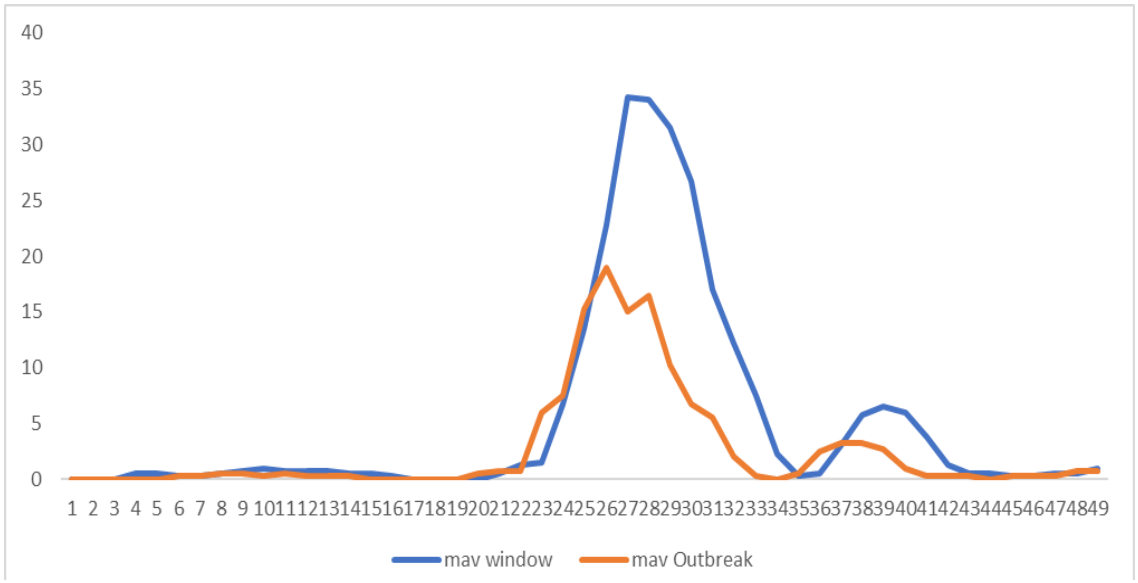

Figure 4.3: Disease prediction smoothed graph using moving average of 4 for EVD tweets in Nigeria

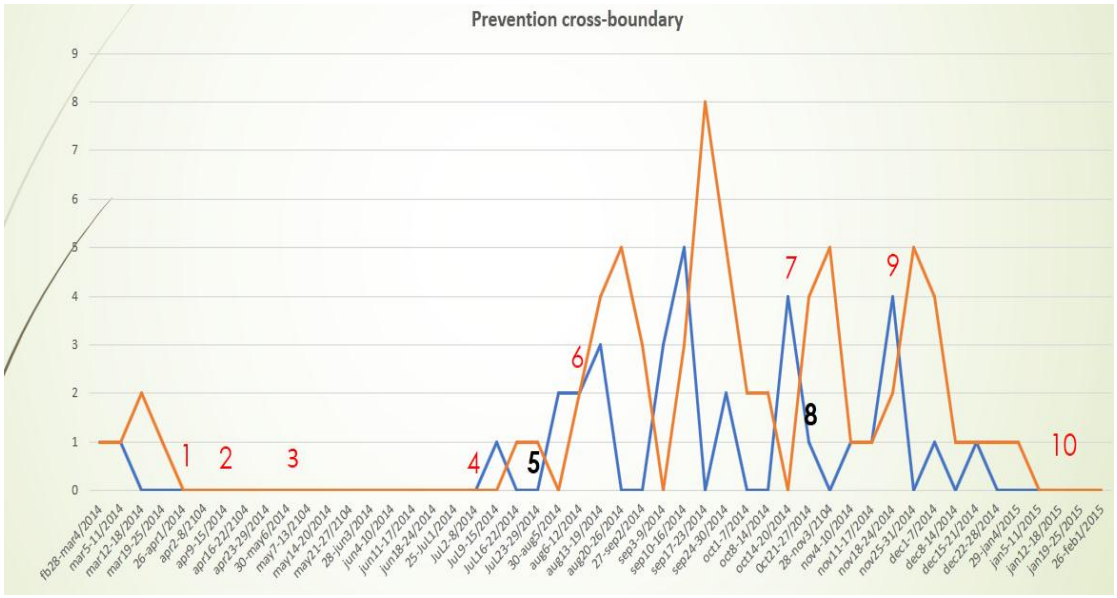

Figure 4.4: Graph of prevention and aggregated window

The data points have a span of seven days: data point 1 is equivalent to data collected from 26 February to 4 March 2014 and data point 49 is equivalent to data collected from 25 January to 31 January.

These points show some consistency with reported cases in Nigeria and its neighbouring countries:

- Nigerian Health Minister confirms first case of Ebola 6 August 2014, five new cases being treated. Source: Mark (2014).
- Ebola strikes at the heart of Nigeria, doctor who treated Sawyer dies 19 August 2014. Source: Thisday (2014).

Some merits of the approach are that with cross boundary, not all spikes in signal level are considered as cases of new outbreaks, which checks alert errors. Our approach is sensitive to both high and low-level changes in signal levels. The decision process is adaptive and varies based on the input levels, without a need to set constant threshold values, and it checks spill-over after the trauma of a new outbreak.

The implementation was done using the following setups:

- Keras and python for the deep-learning model, which used tensorflow backend for the keras library;
- Hadoop file system;
- Apache Nifi for generating streams;
- Apache Solr for information extraction.

CHAPTER 5
SUMMARY AND RECOMMENDATION

5.1 Summary

This work contributes to an emerging field of deep learning for disease control, we applied a character-level approach for text analytics that defeats the need for tasking model augmentation methods used in word vector learning and rule-based procedures. Even with comparatively little data, an NLP model with comparative performance in short text analysis was developed. The disease prediction model was built to check the frailties of some previously failed methods that were implemented for infectious disease monitoring and control; such as, the increase in information search on a particular disease may not translate to an outbreak.

5.2 Recommendation

It will be worthwhile if time and resources are invested to make disease-related short text corpus publicly available to aid research in this area and to implement NLP tasks for named entity recognition (NER) to track the location of the outbreak reports.

Work to identify and control factors which directly or indirectly influence the inexplicable occurrence and reoccurrence of disease outbreaks in developing nations would reduce mortality rate. This will make epidemiology branch out into more fields.

References

Andrychowicz, M., Denil, M., Gomez, S., Hoffman, M. W., Pfau, D., Schaul, T., de Freitas, N. (2016). Learning to learn by gradient descent by gradient descent. Advances in Neural Information Processing Systems 29 (NIPS 2016), 1-17. Retrieved from https://papers.nips.cc/paper/6461-learning-to-learn-by-gradient-descent-by-gradient-descent.pdf

Baars, H., & Kemper, H.-G. (2008). Management support with structured and unstructured Data − an integrated business intelligence framework. Information Systems Management, 25(2), 132–148. https://doi.org/10.1080/10580530801941058

Bansal, S., Chowell, G., Simonsen, L., Vespignani, A., & Viboud, C. (2016). Big data for infectious disease surveillance and modeling. Journal of Infectious Diseases, 214 (Suppl 4), S375–S379. https://doi.org/10.1093/infdis/jiw400

Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A Neural Probabilistic Language Model. The Journal of Machine Learning Research, 3, 1137–1155. https://doi.org/10.1162/153244303322533223

Brauer, F. & Castillo-Chavez, C. (2012). Mathematical models in population biology and epidemiology. DOI: 10.1007/978-1-4614-1686-9

Chan, E. H., Brewer, T. F., Madoff, L. C., Pollack, M. P., Sonricker, A. L., Keller, M., … Brownstein, J. S. (2010). Global capacity for emerging infectious disease detection. Proceedings of the National Academy of Sciences, 107(50), 21701–21706. https://doi.org/10.1073/pnas.1006219107

Choi, J., Cho, Y., Shim, E., & Woo, H. (2016). Web-based infectious disease surveillance systems and public health perspectives: a systematic review. BMC Public Health, 16(1), 1238. https://doi.org/10.1186/s12889-016-3893-0

Chopra, P., & Yadav, S. K. (2017). Restricted Boltzmann machine and softmax regression for fault detection and classification. Complex & Intelligent Systems. https://doi.org/10.1007/s40747-017-0054-8

Culotta, A. (2010). Towards detecting influenza outbreaks by analyzing Twitter messages. Proceedings of the first workshop on social media analytics (SOMA'10), pp.115-122. https://doi.org/10.1145/1964858.1964874

De Brébisson, A., & Vincent, P. (2015). An exploration of Softmax alternatives belonging to the spherical loss family. International Conference on Learning Representations (ICLR 2016) 1–9. Retrieved from http://arxiv.org/abs/1511.05042

Demeester, T., Rocktäschel, T., & Riedel, S. (2016). Lifted rule injection for relation embeddings. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp.1389-1399. Retrieved from http://aclweb.org/anthology/D/D16/D16-1146.pdf

Dowdle, W. R. (1998). The principles of disease elimination and eradication. Bulletin of the World Health Organization, 76(SUPPL. 2), 22–25. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/ PMC2305684/pdf/bullwho00391-0020.pdf

Farzindar, A., & Inkpen, D. (2015). Natural language processing for social media, i, 1 PDF (xix, 146 pages). https://doi.org/10.2200/S00659ED1V01Y201508HLT030

Fuseware, & World Wide Worx. (2014). SA Social Media Landscape 2014, 3–6. Retrieved from http://www.worldwideworx.com/wp-content/uploads/2013/10/Exec-Summary-Social-Media-2014.pdf

Geist, M. (2015). Soft-max boosting. Machine Learning, 100(2–3), 305–332. https://doi.org/10.1007/s10994-015-5491-2

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. AISTATS '11: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, 15, 315–323. https://doi.org/10.1.1.208.6449

Hashimoto, S., Murakami, Y., Taniguchi, K., & Nagai, M. (2000). Detection of epidemics in their early stage through infectious disease surveillance. International Journal of Epidemiology, 29(5), 905–910. https://doi.org/doi: 10.1093/ije/29.5.905

Kanhabua, N., & Nejdl, W. (2013). Understanding the diversity of tweets in the time of outbreaks. Proceedings of the 22nd International Conference on World Wide Web – WWW '13 Companion, 1335–1342. https://doi.org/10.1145/2487788.2488172

Kebede, S., Duales, S., Yokouide, A., & Alemu, W. (2010). Trends of major disease outbreaks in the African region, 2003-2007. East African Journal of Public Health, 7(1), 20–29. https://doi.org/10.4314/eajph.v7i1.64672

Komninos, A., & Manandhar, S. (2016). Dependency based embeddings for sentence classification tasks. Proceedings of NAACL, 1490–1500. Retrieved from http://www.aclweb.org/anthology/N16-1175

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 1–9. Retrieved from https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review, 104(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Langseth, J., Vivatrat, N., & Sohn, G. (2005). Analysis and transformation tools for structured and unstructured data. U.S. Patent Application No. 11/172,957.

Levy, J. P., & Bullinaria, J. A. (2000). Learning lexical properties from word usage patterns: Which context words should be used? Connectionist Models of Learning, Development and Evolution, 273–282. Retrieved from https://link.springer.com/chapter/10.1007/978-1-4471-0281-6_27

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. Computational Linguistics, 19(2), 313–330. Retrieved from https://repository.upenn.edu/cgi/viewcontent.cgi?article=1246&context=cis_reports

Mark, M. (2014, August 6). Ebola outbreak: Nurse who treated first victim in Nigeria dies. The Guardian, Retrieved from https://www.theguardian.com/world/2014/aug/06/ebola-outbreak-nurse-nigeria-dies

Mikolov, T., Corrado, G., Chen, K., & Dean, J. (2013). Efficient estimation of word representations in vector space. Proceedings of the International Conference on Learning Representations (ICLR 2013), 1–12. Retrieved from https://arxiv.org/pdf/1301.3781.pdf

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean J. (2013). Distributed representations of words and phrases and their compositionality. Advances in Neural Information Processing Systems 26 (NIPS, 2013). Retrieved from https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf

Ofoghi, B., Mann, M., & Verspoor, K. (2016). Towards early discovery of salient health threats: A social media emotion classification technique. Pacific Symposium on Biocomputing, 21, 504–515. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26776213

Parke, P. (2013, January 14). How many people use social media in Africa. Retrieved from http://edition.cnn.com/2016/01/13/africa/africa-social-media-consumption/index.html

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Rong, X. (2014). word2vec Parameter Learning Explained, 1–21. Retrieved from https://arxiv.org/abs/1411.2738

Shen, M., Xiao, Y., & Rong, L. (2015). Modeling the effect of comprehensive interventions on Ebola virus transmission. Scientific Reports, 5, 15818. https://doi.org/10.1038/srep15818

Simonsen, L., Gog, J. R., Olson, D., & Viboud, C. (2016). Infectious disease surveillance in the big data era: Towards faster and locally relevant systems. Journal of Infectious Diseases, 214 (Suppl 4), S380–S385. https://doi.org/10.1093/infdis/jiw376

Spinage, C. A., & House, W. (2012). African ecology - Benchmarks and historical perspectives. Heidelberg Dordrecht London New York: Springer. Retrieved from https://www.springer.com/gp/book/9783642228711

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15, 1929–1958. https://doi.org/10.1214/12-AOS1000

Akinremi, A., & Balogun, S. (2014, August 20). Ebola strikes at the heart of Nigeria: Ameyo, daughter of Kwaku Adadevoh, granddaughter of Herbert Macaulay dies. Thisdaylive, Retrieved from https://web.archive.org/web/20140821181052/http://www.thisdaylive.com/articles/ebola-strikes-at-the-heart-of-nigeria-ameyo-daughter-of-kwaku-adadevoh-great-grand-daughter-of-herbert-macaulay-dies/186843/

Tobergte, D. R., & Curtis, S. (2013). Improving neural networks with Dropout. Journal of Chemical Information and Modeling, 53(9), 1689–1699.

Troppy, S., Haney, G., Cocoros, N., Cranston, K., & DeMaria, A. (2014). Infectious disease surveillance in the 21st century: An integrated web-based surveillance and case management system. Public Health Reports (Washington, D.C. : 1974), 129(2), 132–8. https://doi.org/10.1177/003335491412900206

Van Den Driessche, P., & Watmough, J. (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical Biosciences, 180, 29–48. https://doi.org/10.1016/S0025-5564(02)00108-6

Varol, H. A. (2016). MOSES : A Matlab-based open-source stochastic epidemic simulator, 2636–2639. https://doi.org/10.1109/EMBC.2016.7591271

Vijayarani, S., Ilamathi, J., & Nithya, M. (2015). Preprocessing Techniques for Text Mining: An Overview. International Journal of Computer Science & Communication Networks, 5(1), 7–16. Retrieved from http://www.ijcscn.com/Documents/Volumes/vol5issue1/ijcscn2015050102.pdf

Wakefield, T., & Bean, D. (2005). Visualization of integrated structured and unstructured data. U.S. Pa-tent Application No. 10/729,388, 2005.

Xia, Z.-Q., Wang, S.-F., Li, S.-L., Huang, L.-Y., Zhang, W.-Y., Sun, G.-Q., Jin, Z. (2015). Modeling the transmission dynamics of Ebola virus disease in Liberia. Scientific Reports, 5(1), 13857. https://doi.org/10.1038/srep13857

Zhang, X., & LeCun, Y. (2015). Text understanding from scratch. Retrieved from https://arxiv.org/abs/1502.01710

Zhao, L., Chen, J., Chen, F., Wang, W., Lu, C. T., & Ramakrishnan, N. (2016). SimNest: Social media nested epidemic simulation via online semi-supervised deep learning. Proceedings - IEEE International Conference on Data Mining, ICDM, 2016, 639-648. https://doi.org/10.1109/ICDM.2015.39