Journal, Advances in Mathematical & Computational Sciences
Vol. 8 No. 1, March 2020
www.mathematics-computationaljournal.info

Article Citation Format
Enikuomehin, A.O., Odugbesan, I.O., Aiyeniko, O., Onugwu, J. E., Kanu, O.K. (2020): On LSTM Implementation for Emotion Detection using Natural Language Processing (NLP) Toolkit. Journal of Advances in Mathematical & Computational Sc.
Vol. 8, No. 1. Pp 31-42

Article Progression Time Stamps

Article Type: Research Article
Manuscript Received 7th January, 2020
Final Acceptance: 11th March, 2020
Article DOI: dx.doi.org/10.22624/AIMS/MATHS/V8N1P3

On LSTM Implementation for Emotion Detection using Natural Language Processing (NLP) Toolkit

Enikuomehin, A.O*., Odugbesan, I.O., Aiyeniko, O., Onugwu, J. E., Kanu, O.K.,
Department of Computer Science
Lagos State University
Ojo, Lagos, Nigeria.
***E-mail**: Oluwatoyin.enikuomehin@lasu.edu.ng

ABSTRACT

Emotion modeling and it's several corresponding applications continues to draw attention in many domains as these models are now implemented in solving complex man-mind related problems. Emotion is an integral part of human interaction as every individual exhibit a form of emotion at one point or the other. It may be expressed by triad; person's speech, facial expression and written text generally referred to as speech, facial and text based emotions. Reports of existing research confirms that many interpretations of human facial expression, as a determinant of how a person feels, has largely not been correct. Coincidentally, Natural Language Processing (NLP) techniques have been severally reported as an effective method for detecting emotions. The Natural Language Processing (NLP) domain is a tract of Artificial Intelligence and Linguistics research with several successful applications across many domains ranging from text analysis, machine translation, summarization, linguistic evolution and discourse analysis amongst others. This work proposes a technique for effective interpretation of human emotion using NLP toolkit. This will help in many domains including areas of criminal investigations. We identify the seven states of emotions; (neutral, joy, sadness, surprise, anger, fear, disgust) and present a model for proper identification of emotions from facial expression. We identified the elementary subjects describing respective expressions for 16 subjects between the ages of 17 and 25 as contained in the C16 dataset from the Yale Face B database using their mapped features over 9 poses and 64 illumination conditions. The common attributes were calculated using LSTM. This is a special case of RNN. The result was able to validate earlier report of Donahue, J. et al. 2017 with an Accuracy score of 71.66%. This finding can be adapted into the design of real life system for effective placement of emotions as such will improve the performance of institutions like the judiciary where the bench officer uses experience to denote and decode remorsefulness.

Keywords: Emotional modeling, Natural Language Processing, Toolkit, LSTM, RNN

1. INTRODUCTION

The world is ever-changing, new technologies are evolving and computer application is gradually permeating every facets of human endeavour. Since the computer does not understand human language without the involvement of experts, a barrier has been created. The bottleneck is often the inability to have natural (human) communication between computer and user (Sluis, 2008). The advent of Natural Language Processing (NLP) came as a panacea to solve the problem. It is a tract of Artificial Intelligence and Linguistics, devoted to make computers understand the statements or words written in human languages. It is met to ease the user's work and to satisfy the wish to communicate with the computer in natural language (Khurana, Koli, Khatter, & Singh, 2017). The extraction and recognition of the facial expression has been the topic of various research projects aimed at enabling smooth interaction between computers and their users (Tumpi, Rahman, & Ali, 2007)

1.1 Background

Natural language processing (NLP) has recently gained much attention for representing and analyzing human language computationally. It has spread its applications in various fields such as machine translation, email spam detection, information extraction, summarization, medical, and question answering (Khurana et al., 2017). Emotion is an integral part of human interaction as every individual exhibit a form of emotion at one point or the other. It is a strong feeling such as love, hate, or anger. It is a strong feeling deriving from one's circumstances, mood, or relationships with others. Emotion Detection can be seen as an important field of research in human-computer interaction. Emotions may be expressed by a person's speech, face expression and written text known as speech, facial and text based emotion respective (Shivhare & Khethawat, 2012). It is very difficult to predict human emotions quantitatively. Though facial expressions and gestures are the best ways to figure out one's emotions, it becomes difficult to identify them as the age of a person increases, because people learn to control their expressions with age and experience. Moreover, the expressions and gestures reveal only the external emotions like anger, happy and sad but fail to do in case of emotions like disgust, boredom etc. To overcome this, different methods are discovered for emotion recognition (Lalithaa, Geyasrutia, Narayanana, & Shravani, 2015).

Facial expression is one of the most powerful, natural, and immediate means for human beings to communicate their emotions and intentions. The extraction and recognition of the facial expression has been the topic of various research projects aimed at enabling smooth interaction between computers and their users. Automatic detection of facial expressions is important for understanding human emotions and for paralinguistic communication, for designing multimodal user interfaces, and for related applications, such as human identification (Tumpi et al., 2007); hence the research aimed at modeling human facial expression using Long Short-Term Memory (LSTM).

Research has shown than Human emotion is very difficult to determine just by looking at the face and also the behavior of a person, though not impossible. Even at that, the basic human emotion can be detected through facial expression or body movement (Wan, Hanif, & Noraini, 2016). Emotions are said to be derived from the muscles under the facial skin and human face undergoes several states of facial expression in a day. Facial expressions are manifestation of human emotions; they are like a front end application for the emotions of a human (Mohan & Tripathi, 2018). Emotion modeling and its several corresponding applications continues to draw attention in many domains as these models are now implemented in solving complex man-mind related problems. There have been various technological devices designed to help detect the emotion of any individual, one of such is the use of Long-Short Term Memory (LSTM) which is a special case of Recurrent Neural Network (RNN).

The extraction and recognition of facial expression has been the topic of various researches subject to enable smooth interaction between computer and their users (Teo, De Silva, & Vadakkepat, 2004). Recognizing human facial expression and emotion by computer is an interesting and challenging problem. Recently there has been a growing interest in improving the interaction between humans and computers. It is argued that to achieve effective human-computer intelligent interaction, there is a need for the computer to interact naturally with the user, similar to the way humans interact (Azcarate, Hageloh, Sande, & Valenti, 2005).

Other than speech and body language, facial expression is one of the most prominent ways by which humans communicate their feelings to other humans. The task of detection of emotions in real time accurately has been a very arduous task as methods, giving good results are generally computationally exhaustive whereas the methods that has low computation time does not yield good results (Vaish, Gupta, & Rathee, 2019). Reports of existing research confirms that many interpretations of human facial expression, as a determinant of how a person feels, has largely not been correct.

Coincidentally, Natural Language Processing (NLP) techniques have been severally reported as an effective method for detecting emotion. Several methods have been adopted to detect and recognize of human facial expression ranging from mere observation to Machine learning. How then can we effectively model human facial expression with less computational time and high accuracy level? A language can be defined as a set of rules or set of symbol. Symbol are combined and used for conveying information or broadcasting the information. Symbols are tyrannized by the rules (Khurana et al., 2017). Natural Language Processing began in the 1950s as the intersection of Artificial Intelligence and linguistics. It was originally distinct from text Information Retrieval (IR), which employs highly scalable statistics-based techniques to index and search large volumes of text efficiently. Currently, NLP borrows from several, very diverse fields, requiring today's NLP researchers and developers to broaden their mental knowledge-base significantly (Nadkarni, Ohno-Machado, & Chapman, 2011).

Traditional methods assume that natural language may be described as a set of components that may be combined into progressively more complex components, with the components being understood to be based on conventional notions about language. These models assume that terms fall into specific categories, or parts-of-speech, and that terms are combined using grammatical rules to produce sentences (Losee, 2001)Suppose you received an email from a friend that read: "Quieres ir al cine a ver esa película nueva que Jenny dijo que parecía ser muy buena?" You realize immediately that this e-mail is not written in English. If you don't speak the language of the e-mail, you may want to know what the language is, want a literal translation, or maybe want just the gist of the e-mail (do you want to go to the movies?).

The computational processes involved in this type of task are collectively referred to as NLP, where the term "natural language" applies to the human language content of what is being processed by the computer. NLP is concerned with computer algorithms that analyze, modify, augment, or generate human language and methods that range from assigning probabilities to words or sequences of words (e.g., word prediction or completion) to full-scale transformation of sentences into new sentences (e.g., sentence simplification) (Higginbotham, Lesher, Moulton, & Roark, 2012).

2. NLP AND EMOTIONS

Natural Language Processing (NLP) is a field at the intersection of computer science, artificial intelligence, and linguistics. The goal is for computers to process or "understand" natural language in order to perform tasks like Language Translation and Question Answering. With the rise of voice interfaces and chatbots, NLP is one of the most important technologies of the information age a crucial part of artificial intelligence (James, 2018).

NLP is a tract of Artificial Intelligence and Linguistics, devoted to make computers understand the statements or words written in human languages. Natural language processing came into existence to ease the user's work and to satisfy the wish to communicate with the computer in natural language. Since all the users may not be well-versed in machine specific language, NLP caters those users who do not have enough time to learn new languages or get perfection in it (Khurana et al., 2017). Natural Language Processing basically can be classified into two parts i.e. Natural Language Understanding and Natural Language Generation which evolves the task to understand and generate the text (Khurana et al., 2017).

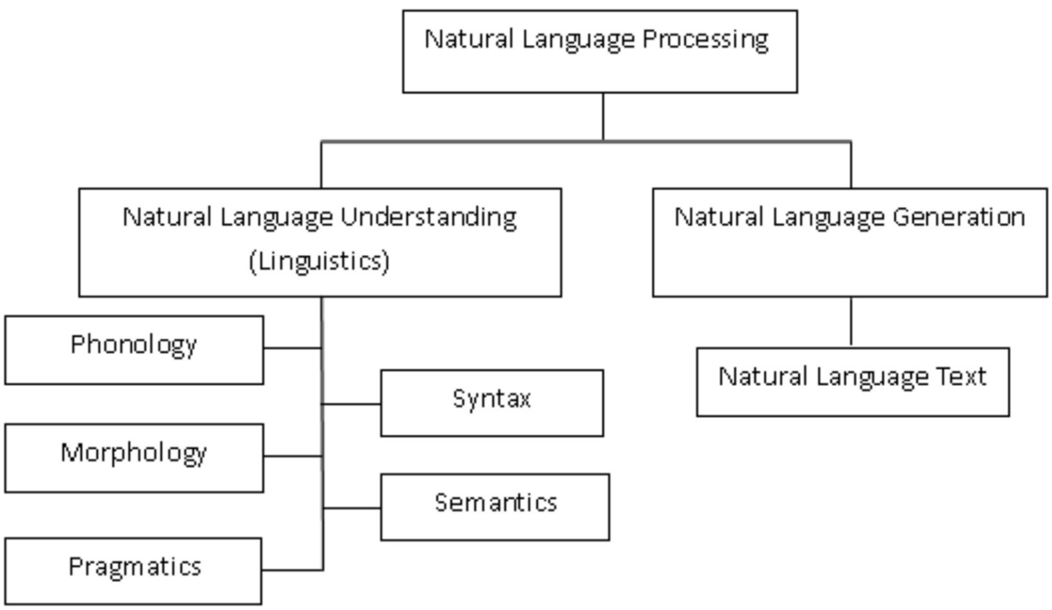


Figure 1: Broad Classification of NLP

Linguistics is the science of language which includes Phonology that refers to sound, Morphology word formation, Syntax sentence structure, Semantics syntax and Pragmatics which refers to understanding (Khurana et al., 2017). Most NLP systems process input via statistical language models trained on observations of natural language using machine learning techniques. Most of these NLP technologies are powered by Deep Learning — a subfield of machine learning. Deep Learning provides a very flexible, universal, and learnable framework for representing the world, for both visual and linguistic information. Initially, it resulted in breakthroughs in fields such as speech recognition and computer vision. Recently, deep learning approaches have obtained very high performance across many different NLP tasks. These models can often be trained with a single end-to-end model and do not require traditional, task-specific feature engineering (James, 2018).

Among the many applications making core use of NLP algorithms are automatic machine translation from one language to another, extraction of structured information from large language corpora, speech recognition, and processing of spoken language. Other areas such as dysarthric speech recognition, sentence simplification, context input, and brain computer interfaces are being actively researched (Higginbotham et al., 2012). There's a fast-growing collection of useful applications derived from this field of study. They range from simple spell checking, keyword search, finding synonyms; extracting information from websites such as: product price, dates, location, people, or company names; classifying: reading level of school texts, positive/negative sentiment of longer documents; machine translation; spoken dialog systems; complex question answering.

These applications have been used abundantly in industry: from **search** (written and spoken) to online advertisement **matching**; from automated/assisted **translation** to **sentiment analysis** for marketing or finance/trading; and from **speech recognition** to **chatbots/dialog agents** (automating customer support, controlling devices, ordering goods) (James, 2018).Most of these NLP technologies are powered by Deep Learning - a subfield of machine learning. Deep Learning provides a very flexible, universal, and learnable framework for representing the world, for both visual and linguistic information. Initially, it resulted in breakthroughs in fields such as speech recognition and computer vision. Recently, deep learning approaches have obtained very high performance across many different NLP tasks. These models can often be trained with a single end-to-end model and do not require traditional, task-specific feature engineering.

The beginning of modern AI can be traced to classical philosophers' attempts to describe human thinking as a symbolic system. But the field of AI wasn't formally founded until 1956, at a conference at Dartmouth College, in Hanover, New Hampshire, where the term "Artificial Intelligence" was coined (Lewis, 2014). Since the invention of computers or machines, their capability to perform various tasks went on growing exponentially. Humans have developed the power of computer systems in terms of their diverse working domains, their increasing speed, and reducing size with respect to time. A branch of Computer Science named Artificial Intelligence pursues creating the computers or machines as intelligent as human beings (TutorialPoint, 2015). Intelligence is the computational part of the ability to achieve goals in the world (Kask, 2015). Intelligence is the capacity to learn and solve problems". In particular, it is the ability to solve novel problems, to act rationally and to act like humans.

According to the father of Artificial Intelligence John McCarthy, it is *"the science and engineering of making intelligent machines, especially intelligent computer programs".* Artificial Intelligence is a way of making a computer, a computer-controlled robot, or a software think intelligently, in the similar manner the intelligent humans think. AI is accomplished by studying how human brain thinks, and how humans learn, decide, and work while trying to solve a problem, and then using the outcomes of this study as a basis of developing intelligent software and systems (TutorialPoint, 2015).
It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable (Kask, 2015). Figure 2 is a pictorial representation of the stages of development in Artificial Intelligence:

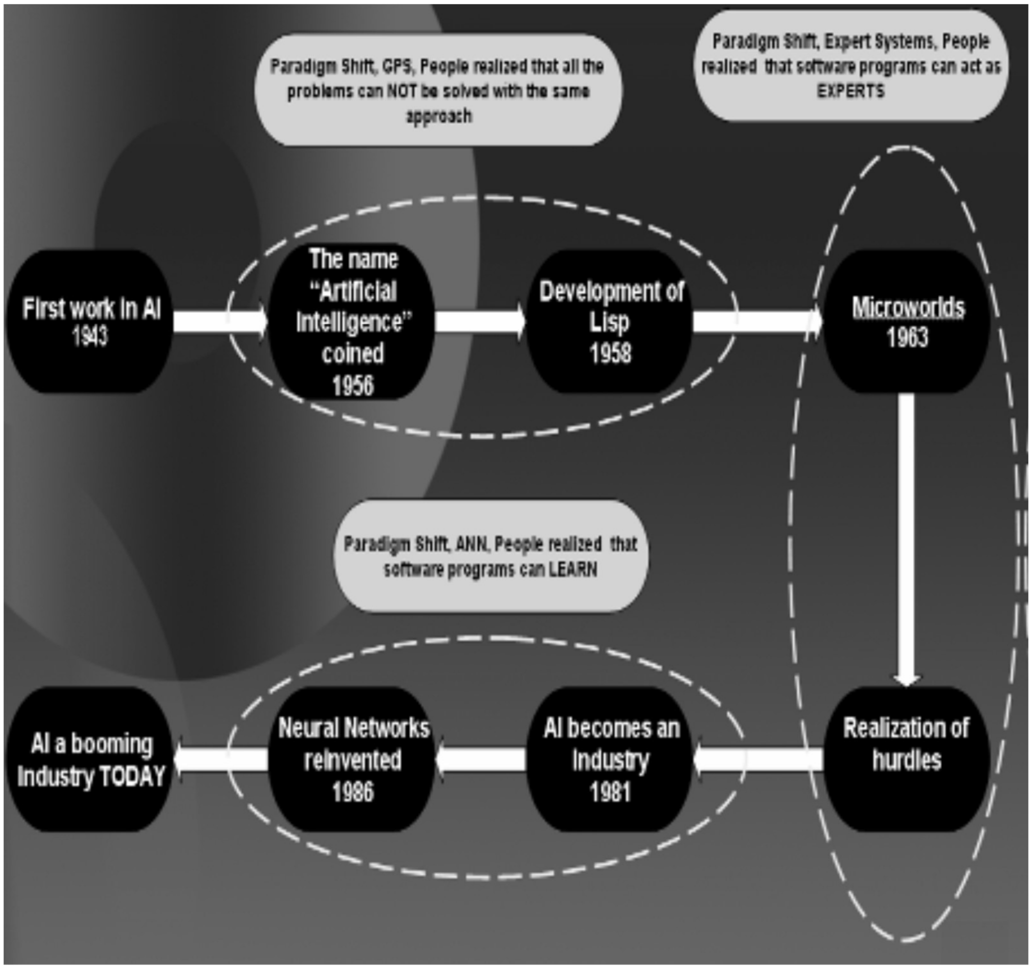


Figure 2: The diagram above summarizes the history and evolution of AI in a comprehensive shape (Alvi, 2018)

Emotions are a common term used daily. It plays an important role in human interactions as they let people articulate themselves without words. Emotions include cognitive appraisal, bodily language, action tendencies, expressions, and feelings, People would not be able to get along with each other without emotions (Tzuu-Hseng, Ping-Huan, Ting-Nan, & Po-Chien, 2017). Emotions are the way of expression of the person's feelings which has an high influence on the decision making tasks. Emotions are the combinations of the feelings, behavior, physiology, conceptualization and experience that are expressed by any living beings. The emotions can be expressed through facial expressions, text representation and through speech (Naveenkumar, Vinayakumar, & Soman, 2019). Although there are various definitions of "emotion", the commonly used term is based on the psychology term where, emotion is defined as "a complex state of feeling that results in physical and behaviour". Emotion plays an essential role in rational decision-making, perception, learning and a variety of functions".

The implicit communication between a communicator plays a significant role in human social interactions (Latif, Yusof, Sidek, N., & Sado, 2016). It is a communication method for describing inner feelings through the physical and real world in the form of body language involves facial expressions and body movement. Emotion is synonymous to all mankind (Wan et al., 2016). There are two main categories on theory of emotion; cognition and somatic. The cognition category appeal to a necessary element of emotion and the subjective manifestation which could be conscious or unconscious, intentional or unintentional and take a form of a judgment or a thought. The somatic category on the other hand relies on somatic features and seeks to describe emotional expressions and perceptions of emotional expressions. This theory believes that body responses trigger emotional reactions (Latif et al., 2016).

Various studies have been carried out to determine the various kinds of emotions, in a previous study conducted by D'Mello et al., eight emotions were identified: anger, boredom, confusion, contempt, curiosity, disgust, eureka, and frustration. Anger was defined as a strong feeling of displeasure and usually of antagonism. Boredom was defined as the state of being weary and restless through lack of interest. Confusion was defined as a failure to differentiate similar or related ideas. Contempt was defined as the act of despising, a lack of respect or reverence for something. Curious was defined as an active desire to learn or to know. Disgust was defined as marked aversion aroused by something highly distasteful. Eureka was defined as a feeling used to express triumph on a discovery. Frustration was defined as making vain or ineffectual efforts however vigorous; a deep chronic sense or state of insecurity and dissatisfaction arising from unresolved problems or unfulfilled needs (Heraz & Frasson, 2011). Emotions have been interpreted in different fields; in the field of physiology, emotion was the result of the reaction of the body while in theory the field of neurological state, activity in the brain is the source that generates the emotional response.

Meanwhile, in the field of the cognitive theory states that mental activity is a very important role in the formation of emotion. Emotions can generally be categorized into three major classifications: Basic (happy, sad, fear, anger, surprise, and disgust), motivation (thirst, hunger, pain, and mood) and the conscious self / Social (shame, dignity, and guilt) (Wan et al., 2016). It is very difficult to predict human emotions quantitatively. Though facial expressions and gestures are the best ways to figure out one's emotions, it becomes difficult to identify them as the age of a person increases, because people learn to control their expressions with age and experience. Moreover, the expressions and gestures reveal only the external emotions like anger, happy and sad but fail to do in case of emotions like disgust, boredom etc (Lalithaa et al., 2015). A reliable estimate of user's emotion in a particular scenario is valuable information for any affective computing system especially if it can be acquired automatically and in real time (Matlovic, Gaspar, Moro, Simko, & Bielikova, 2016).

Emotion recognition involves considerable information including facial expressions, body language, pitch and tone of voice, and semantics. Facial expressions are produced by the contraction and relaxation of some facial muscles. It expressions are crucial because they convey considerable information that can be widely used in various applications in different fields. Facial expressions can convey the same information across different cultures and countries. Long and Short Term Memory (LSTM), an enhanced recurrent neural network (RNN), is used to capture the temporal and contextual information of facial expressions (Tzuu-Hseng et al., 2017). Long-short term memory network, or abbreviating as LSTM, is one of most popular recurrent neural network structure in deep learning field. Invented by Schmidhuber in 1997, LSTM avoids the vanishing gradient issue by adding three gated units: forget gate, input and output gates, through which the memory of past states can be efficiently controlled. LSTM is widely used in many areas, mostly in machine learning application field, including speech recognition, natural language processing and other pattern recognition applications (Wang, 2017).

Long short-term memory (LSTM) recurrent neural network is used to determine the relationship between the transformation of facial expressions in image sequences and the six basic emotions. LSTM networks try to combat the vanishing/exploding gradient problem by introducing gates and an explicitly defined memory cell. Each neuron has a memory cell and three gates: input, output and forget. The function of these gates is to safeguard the information by stopping or allowing the flow of it. LSTM's has shown to be able to learn complex sequences, and are currently very hip and have been used a lot in machine translation. It is the default model for sequence labeling tasks, which have lots and lots of data (James, 2018).

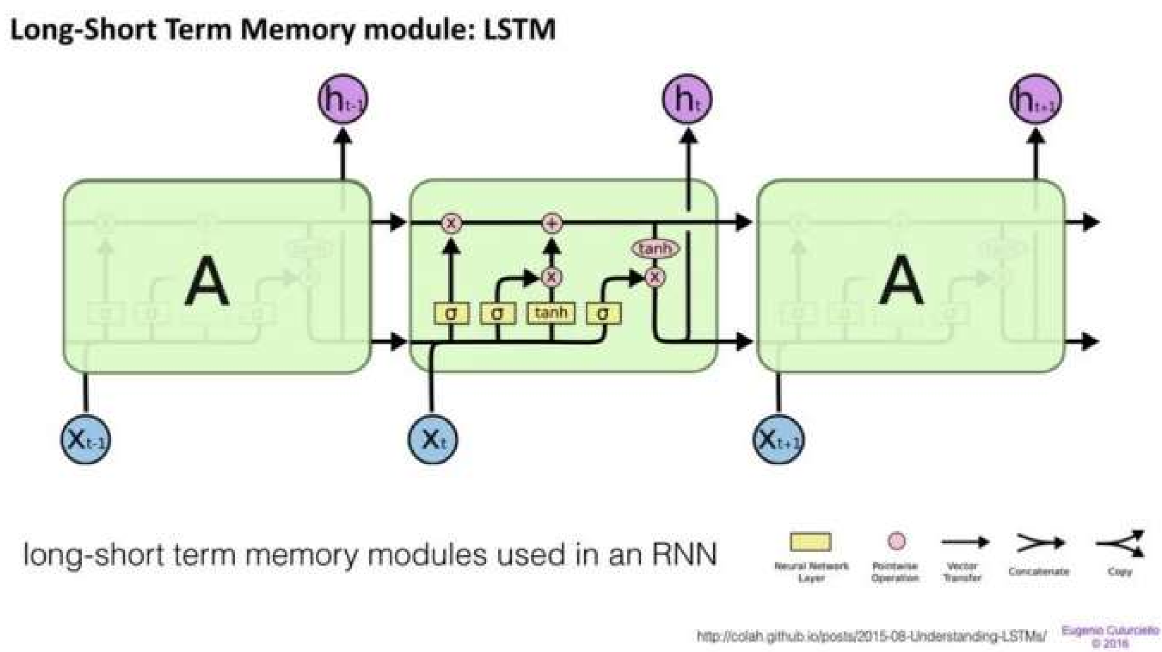


Fig 3: Long-Short Term Memory Module

Research has shown that many deep learning methods such as CNN, RNN, improved deep neural networks, and enhanced models LSTM have been used to solve some difficult tasks and improve performance (Tzuu-Hseng et al., 2017).

3. METHOD

Emotion recognition is a complex task that even some humans have a hard time recognizing certain emotions. Such is the case of blocked and/or micro emotions, where a skilled person can conceal an emotion that is being expressed with facial expressions. Unfortunately, common people cannot accurately recognize real and fake facial expressions. This, in machine learning, can be described as a classification problem as well as a supervised machine learning problem.Here, materials used in terms of data collection and the Recurrent Neural Network based LSTM method used in the identification of emotion from the data are presented. We chose RNN-based LSTM because deep learning has produced acceptable results in the area of computer vision in the identification of human expression as well as postures and in other areas of image and pattern recognition

We identify the seven states of emotions; (neutral, joy, sadness, surprise, anger, fear, disgust) and present a model for proper identification of emotions from facial expression. The architecture of the emotion recognizer is given in Figure 3.1. It is divided into three phases: the Data Identification and Collection, Model Building and Model Training.In the first phase, we identified the data required for this work and sourced for it. The collected data is then processed by performing the one-hot encoding labels. The second phase is the building of the LSTM model. This phase includes the construction of the LSTM model, addition of activation, construction of input and output architectures and parameter tuning. The third phase is for the model building and testing. This involves the systematic chunking of the dataset (output of stage 1) into training and testing. The model is trained and its performance will be evaluated.

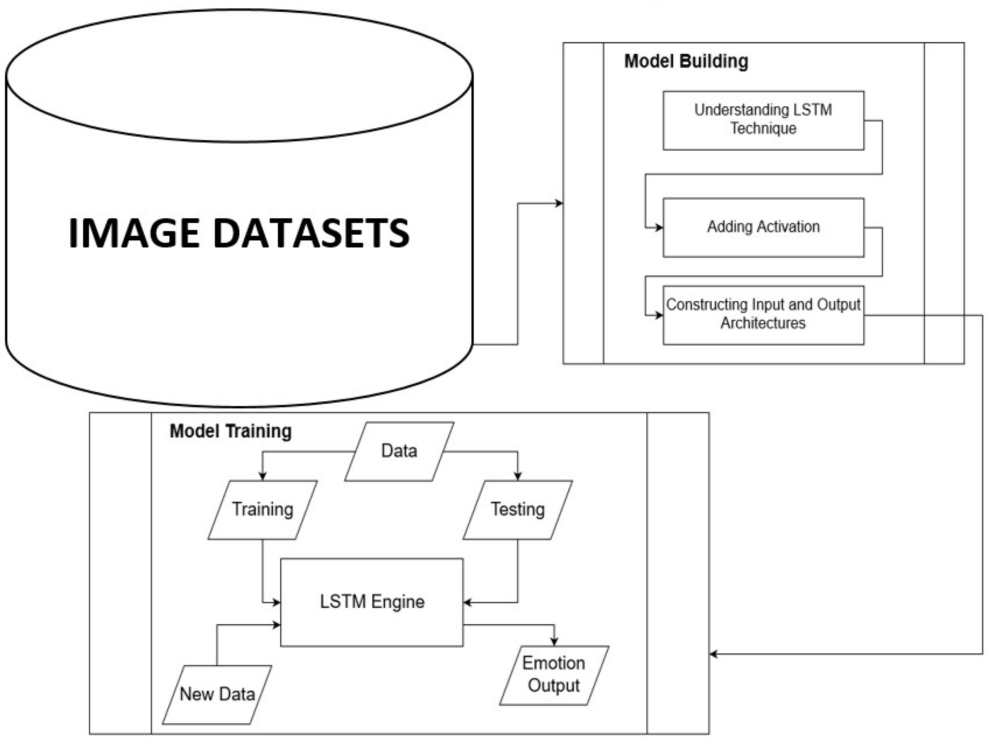


Figure 3.1 System Architecture of the Proposed LSTM Emotion Recognizer

The LSTM Model is concerned with the building of the three layers (first layer, hidden layer and dense layer). This is followed by adding the activation functions and, lastly, constructing the I/O architectures. This is further viewed from Figure 3.2.

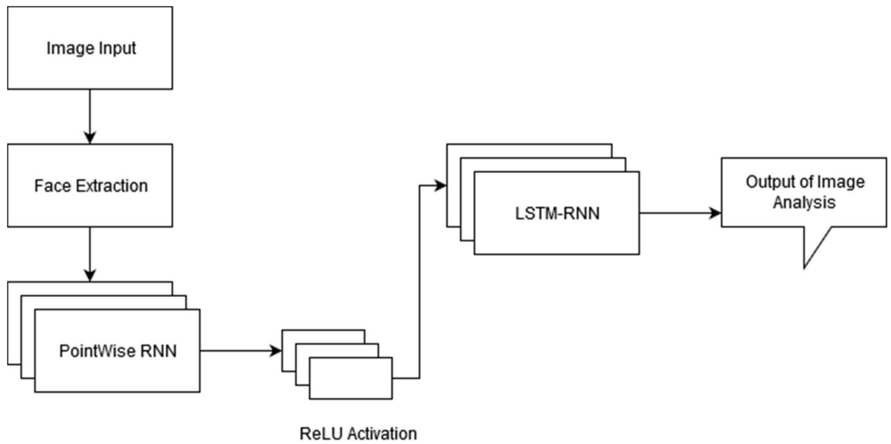
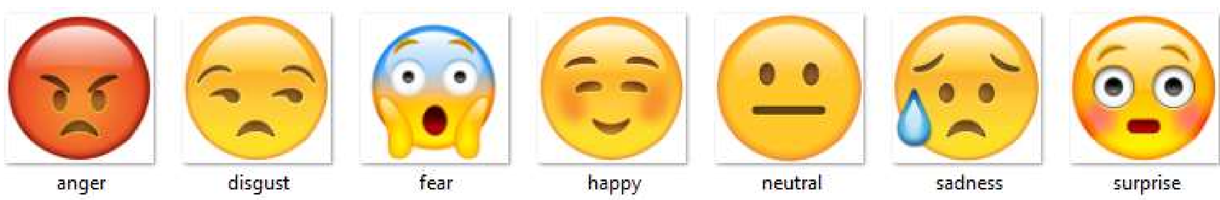


Figure 3.2 Model Workflow

We identified the elementary subjects describing respective expressions for 16 subjects between the ages of 17 and 25 as contained in the C16 dataset from the Yale Face B database using their mapped features over 9 poses and 64 illumination conditions.



The common attributes were calculated using LSTM. This is a special case of RNN. Long Short Term Memory networks (LSTM) solved the problem of long-term dependencies. These special kind of RNNs are capable of learning both short-term and long-term dependencies although they are specialised for long-term relationships. The emotion recognition model goes through the forget, input, update and output gates.This is extracted from Figure 3.3.

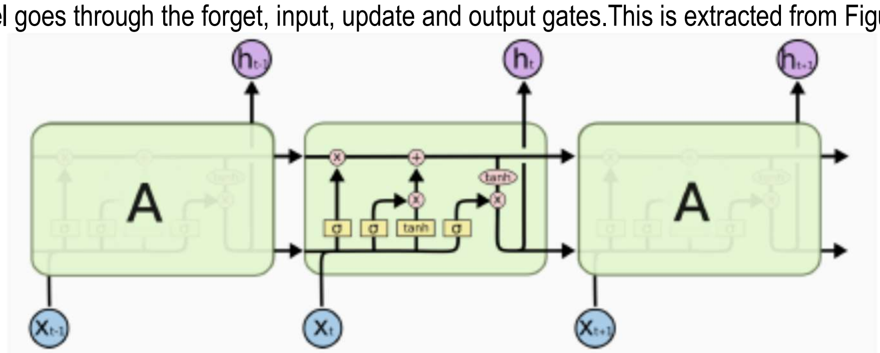


Figure 3.3 The LSTM Architecture

4. RESULT

The result was able to validate earlier report of Donahue, J. et al. 2017 with an accuracy score of 71.66%.

5. CONCLUSION

This study will help in many domains including areas of criminal investigations. This finding can be adapted into the design of real life system for effective placement of emotions as such will improve the performance of institutions like the judiciary where the bench officer uses experience to denote and decode remorseless.

REFERENCES

1. Alvi, Z. M. (2018). Artificial Intelligence. 1 - 208.
2. Azcarate, A., Hageloh, F., Sande, K. V. D., & Valenti, R. (2005). Automatic facial emotion recognition. 1 - 10.
3. Heraz, A., & Frasson, C. (2011). Towards a Brain-sensitive Intelligent Tutoring System: Detecting Emotions from Brainwaves. *Advances in Artificial Intelligence, 2011*, 1-13.
4. Higginbotham, J., Lesher, W. G., Moulton, B. J., & Roark, B. (2012). The Application of Natural Language Processing to Augmentative and Alternative Communication. *Article in Assistive technology: the official journal of RESNA*, 12-25.
5. James, L. (2018). The 7 NLP Techniques That Will Change How You Communicate in the Future (Part I).
6. Kask, K. (2015). Introduction to Artificial Intelligence. 1-50.
7. Khurana, D., Koli, A., Khatter, K., & Singh, S. (2017). Natural Language Processing: State of The Art, Current Trends and Challenges. 1 - 25.
8. Lalithaa, S., Geyasrutia, D., Narayanana, R., & Shravani, M. (2015). Emotion Detection using MFCC and Cepstrum Features. *4th International Conference on Eco-friendly Computing and Communication Systems*, 29 - 35.
9. Latif, M. H., Md, Yusof, H., Sidek, S. N., N., R., & Sado, F. (2016). Emotion detection from thermal facial imprint based on GLCM features. *ARPN Journal of Engineering and Applied Sciences., 11*(1), 345-350.
10. Lewis, T. (2014). A brief history of Artificial Intelligence.
11. Losee, R. M. (2001). Natural Language Processing In Support of Decision-Making: Phrases and Part-of-Speech Tagging. *Information Processing & Management, 37 (6)*, 1-19.
12. Matlovic, T., Gaspar, P., Moro, R., Simko, J., & Bielikova, M. (2016). Emotions detection using facial expressions recognition and EEG.
13. Mohan, Y., & Tripathi, V. (2018). Comparative analysis of Facial Expression Detection Techniques Based on Neural Network. *International Journal of Engineering & Technology, 7*(4), 866 - 870.
14. Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Article in Journal of the American Medical Informatics Association*.
15. Naveenkumar, K. S., Vinayakumar, R., & Soman, K. (2019). Emotion Detection using Data Driven Models. *Center for Computational Engineering and Networking (CEN), Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India*, 1-11.
16. Shivhare, S. N., & Khethawat, S. (2012). Emotion detection from text. *Computer Science & Information Technology., 2*, 1 - 7.
17. Sluis, I. V. (2008). Towards Affective Natural Language Generation: Empirical Investigations. *AISB 2008 Convention Communication, Interaction and Social Intelligence, 1st-4th April 2008, University of Aberdeen, Volume 2: Proceedings of the AISB 2008 Symposium on Affective Language in Human and Machine*.

18. Teo, W. K., De Silva, L. C., & Vadakkepat, P. (2004). Facial expression detection and recognition system. *Journal of The Institution of Engineers, Singapore, 44*(3), 14-26.
19. Tumpi, P. N., Rahman, W. R., & Ali, M. H. (2007). An efficient facial expression detection system. *Article in Machine, Graphics and Vision, 16*(3), 377-399.
20. TutorialPoint. (2015). Artificial Intelligence - Intelligence Systems. 1 - 68.
21. Tzuu-Hseng, S. L., Ping-Huan, K., Ting-Nan, T., & Po-Chien, L. (2017). CNN and LSTM Based Facial Expression Analysis Model for a Humanoid Robot.
22. Vaish, A., Gupta, S. D., & Rathee, N. (2019). Enhancing Emotion Detection using Metric Learning Approach. *10.1007-978-981-10-8201-6_36*.
23. Wan, W. O., Hanif, M., & Noraini, H. (2016). Human Emotion detection via brain waves study using Electroencephalogram (EEG). *International journal on Advanced Science, Engineering and Information Technology*.
24. Wang, Y. (2017). A new concept using LSTM Neural Networks for dynamic system identification., 1-8.