This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3308967

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

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Multimodal Deep Learning Crime Prediction Using Crime and Tweets

Sakirin Tam[1], and Ömer ÖzgürTanrıöver[2]

[1] Information Communication Technology Department, University of Puthisastra, Phnom Penh 120204, Cambodia
[2] Computer Engineering Department, Ankara University, Ankara 06830, Turkey

Corresponding author: Sakirin Tam (kirin.it@gmail.com).

ABSTRACT Crime prevention relies on crime prediction as a crucial method to determine the most effective patrol strategy for law enforcement agencies. Various approaches and solutions have been utilized to predict criminal activity. Nonetheless the environment and nature of information for crime prediction is constantly changing. Although, potentially a useful source for gathering sentiments, social media content has been ignored by the prediction models. The utilization of social media for sharing information and ideas has experienced a significant surge. Twitter, in particular, is regarded as a valuable platform for gathering public sentiments, emotions, perspectives, and feedback. In this regard, techniques for analyzing the sentiment of tweets on Twitter have been developed to ascertain whether the textual content conveys a positive or negative viewpoint on crime incident. Therefore, our interest lies in investigating the potential and advantages of fusing the information of sentiment and crime modalities. In this paper, ConvBiLSTM is applied to train the model, features of both tweet and crime modalities were extracted independently at vector level and fused into a single representation that captures the information from all modalities. This study involved collecting and conducting experiments using two datasets. The first dataset consisted of crime incident data obtained from the Chicago police department, specifically covering the period between September 1 and September 30, 2019. The second dataset comprised tweets containing crime-related terminology specific to Chicago. The crime prediction using multimodal data fusion on ConvBiLSTM outperform against other models with 97.75% of accuracy.

INDEX TERMS Multimodal, Data Fusion, Sentiment Analysis, Crime Prediction, Neural Network

I. INTRODUCTION

Crime detection and prediction have emerged as significant and crucial practices in crime analysis, aiming to determine the most effective patrol strategy for law enforcement agencies. Numerous researchers have explored various techniques and solutions utilizing data mining methods to analyze crime. These studies have the potential to streamline and automate the crime analysis process, resulting in increased efficiency and accuracy.

Nevertheless, the nature of crime patterns is dynamic rather than static, consistently evolving and expanding [1]. Social media platforms facilitate public discussions and postings, generating textual data that contains contextual information about users' daily activities. These unstructured posts present valuable data that can be leveraged for crime prediction. Sentiment analysis, also known as opinion mining, employs text analysis and computational linguistic techniques within the field of Natural Language Processing (NLP). Its objective is to identify, extract, and classify subjective information from unstructured text [2]. The primary goal of sentiment analysis is to determine the polarity of sentences by extracting word clues from the sentence context [3]-[5]. Hence, sentiment analysis has gained significant recognition as a valuable technique for extracting meaningful insights from unstructured data sources, such as tweets or reviews. In the business domain, companies utilize sentiment analysis to gain an understanding of customer feedback regarding their products or services. Similarly, in politics, sentiment analysis serves as a decision-making tool to examine public reactions to political events. Various social media platforms, including Twitter, Facebook, Instagram, blogs, reviews, and news websites, enable individuals to widely share their opinions and reviews. The number of Twitter users has grown from 140 million in 2012 to 353.9 million active users in 2023, with approximately 237 million daily active twitter users [6]. These tweets harbor valuable hidden information that can be utilized to ascertain the author's attitude and contextual polarity within the text [7], [8]

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3308967

Crime prediction utilizing sentiment analysis has emerged as an effective approach in recent years [9]. By leveraging sentiment analysis techniques on various data sources such as social media posts, news articles, or online reviews, valuable insights can be derived regarding public attitudes, emotions, and opinions related to crime. Sentiment analysis enables the identification and classification of text as expressing positive, negative, or neutral sentiment, providing a deeper understanding of the prevailing sentiments associated with criminal activities or specific locations. Integrating sentiment analysis into crime prediction models allows law enforcement agencies to gauge the public's perception of safety and allocate resources more effectively. Additionally, sentiment analysis can assist in early detection of emerging crime trends, enabling proactive measures to be taken for crime prevention and ensuring the safety and security of communities.

On the other hand, data of a particular phenomenon or system can be derived from various tools, measurement techniques, experimental setups, and other sources. Given the diverse characteristics of societal processes and environments, it is rare that a single data acquisition method could offer a comprehensive understanding of the phenomenon. As multiple datasets, obtained through different acquisition methods, become more accessible, new possibilities arise, leading to questions that go beyond analysing each dataset independently [10]. Data fusion refers to the merging of data from various modalities offering distinct perspectives on a shared phenomenon, is employed to address inference problems. It holds the potential to resolve such problems with fewer errors compared to unimodal approaches [11]. Data fusion provides several advantages, including complementary, redundant, and cooperative features [12, 13].

In this study, we aim to overcome the drawbacks of the studies found in crime model by developing deep learning multimodal using real time tweets and crime data for a period of one month. We apply our sentiment base deep learning model that we have developed in an earlier study called ConvBiLSTM [14]. The word feature from tweets and crime modalities are extracted independently at vector level and fused into a single representation that captures the information from all modalities. The strength of ConvBiLSTM is that it provides extra training by traversing the data twice from left to right and right to left, there by extracting the vector of the words in context of the information preceding and succeeding it and therefore can capture long term contextual dependencies and global features from the sequential data.

The structure of this paper is as follows: Section 2 provides literature review of relevant study. Details of propose architecture of the model are explained in Section 3. Result and discussion is explained in Section 4. Finally, the possible research improvement to the study and conclusion is provided in Section 5.

II. LITERATURE REVIEW

This section introduces literature on existing study including multimodal data fusion, crime modalities, sentiment modalities, and deep learning model.

A. Multimodal Data Fusion

Multimodal data fusion has a rich research history that can be traced back to audio-visual speech recognition, which was inspired by the well-known McGurk effect [15]. Over time, researchers from various communities, including speech recognition, multimedia content indexing and retrieval, and multimodal interaction, have proposed numerous methods [16-20]. However, due to the limited model capacity of traditional approaches, the popularity of research in multimodal data fusion experienced a lull for a certain period. In recent times, the emergence of deep learning techniques has revitalized the field of multimodal data fusion, presenting new opportunities and avenues for exploration [21].

The methods of multimodal data fusion can be broadly categorized into three groups based on the level of fusion, namely pixel-level fusion, feature-level fusion, and decision-level fusion [22]. Some methods combine elements from these categories. Pixel-level fusion involves directly processing the original multimodal data without performing feature extraction, such as adding pixel values from video inter-frame difference images and original audio waveform diagrams [23]. Pixel-level fusion is a general yet coarse data fusion approach. However, it is seldom used as a standalone method in research models due to its challenging nature in calculating and discovering meaningful information and relationships. Moreover, pixel-level fusion suffers from poor scalability when dealing with high-dimensional multimodal data.

Feature-level fusion involves integrating features immediately after their extraction, often utilizing techniques like principal component analysis (PCA) and linear discriminant analysis (LDA) to reduce the dimensionality of the feature set. For instance, Donahue et al. [24] incorporated a Long Short-Term Memory (LSTM) neural network on top of a convolutional neural network to combine temporal and spatial information in videos. Wu et al. [25] introduced a semantic consistency classification loss function in an early fusion architecture to handle semantic conflicts between modalities, resulting in improved performance. Dai et al. [26] proposed CADNNs, an architecture that embeds prior knowledge into deep neural networks, and applied it to a multimodal deep architecture for traditional Chinese medicine diagnosis [27].

On the other hand, decision-level fusion, which represents the highest level of fusion, combines information from each modality after individual decisions have been made.

Decision-level fusion employs simple fusion mechanisms such as averaging [28], weighting [29], or voting schemes [30] to calculate synthetic values for each modality's decision. Although the fusion process at the decision level may appear straightforward, it can lead to the loss of low-level intermodal information.

B. Crime Modalities

Recent research endeavours have sought to incorporate Twitter data into predictive models for crime prediction. The objective behind integrating Twitter data for crime prediction is to leverage the wealth of information available on the platform regarding users' social behaviours. Geber [31] is credited as the pioneer in incorporating social media content to model crime prediction. Geber utilized Latent Dirichlet Allocation (LDA) on tweets to explore the relationship between tweet content and crime patterns in specific locations. The results showed an improvement compared to models solely relying on traditional historical crime predictors for stalking, criminal damage, and gambling. However, Geber's use of LDA, an unsupervised learning technique, presents challenges as the correlations between word clusters and crimes are not driven by pre-existing theoretical insights. Consequently, the correlations generated may appear relatively insignificant.

Wang et al. [32] employed a novel approach by extracting event-based topics from real-time tweets to predict hit-and-run incidents in Virginia. However, their data source was limited to a manually selected set of news portals, neglecting the vast amount of information contributed by citizens.

Chen et al. [33] incorporated sentiment analysis of tweets along with weather data using Kernel Density Estimation (KDE) to predict thief occurrences in terms of time and location. Nevertheless, their study was constrained to spatial information such as weather data for specific timeframes and locations. Moreover, it was concluded that KDE, being a location-dependent technique, cannot be easily generalized as certain types of crimes may not follow patterns established by previous incidents, and the population of an area can frequently change.

Brandt et al. [34] explored the correlation between mobile populations captured through Twitter's geotagging feature and the occurrence of various types of crime. They found that the absence of tweets was indicative of assaults and thefts. However, these studies primarily focused on geolocation data, disregarding the textual content of tweets.

Similarly, Malleson et al. [35] employed geographic analysis methods to model crime risk using tweets from mobile populations. However, the drawback of these studies was the lack of consideration given to tweet text, as the focus was solely on geolocation data.

Zainuddin et al. [36] implemented sentiment analysis on crime-related tweets using a model based on Natural Language Processing techniques and SentiWordNet. This model had the capability to detect the subjectivity of crime and predict crime based on the presence of hate tweets.

Pang et al. [37] conducted a comparative study utilizing algorithms such as Naïve Bayes, Support Vector Machine, and Maximum Entropy to determine sentiment polarity in movie reviews. These studies proved effective but overlooked the semantic aspect, failing to capture the meaning embedded within the tweets.

Based on previous researches, there are two assumptions can be made; first the publicly available data from Twitter do include features that can portray a correlation between crime pattern predicted from tweets and the actual crime incidents reported. Second, estimation models including social media variables will increase the amount of crime variance explained compared to models that include ‘offline’ variables alone.

C. Deep Learning

Deep learning algorithms have achieved remarkable results in natural language processing area. They represent data in multiple and successive layers. They have the ability to capture the syntactic features from sentences automatically without extra feature extracting techniques, which consume more resource and time. This is the reason why deep learning models have attracted attention from NLP researchers to explore sentiment classification.

By making use of a multi-layer perceptron structure in deep learning, CNN can learn high-dimensional, non-linear, and complex classification. As a result, CNN is used in many applications such as computer vision, image processing, and speech recognition [38], [39]. Kalchbrenner and Blunsom [40] designed Dynamic Convolution Neural Network (DCNN) model for text processing. Kim et al. [41] proposed English text classification by taking word vectors as input into CNN to get sentence-level classification. Even though CNN achieves good results in text classification, it mainly focuses on extracting local features and pays no attention to the context of words, which have much impact on the performance of text classification results [42], [43]. From this motivation of work, an integrated model of CNN with Bi-LSTM was proposed.

Automating the learning and expressing features in neural network enables RNN to integrate the adjacent location of information in NLP effectively. Long Short-Term Memory (LSTM) is one of RNN models [44] that can build a large-scale structure of neural network. LSTM makes good use of memory to avoid gradient problems in RNN [45]. In contrast to CNN and LSTM, RNN pays more attention to context of feature information and can fit into non-linear relations while retaining the sequential of text information [46], [47]. Also, Bidirectional RNN is another type of neural network models that is widely used in text classification [48]. Bidirectional RNN works as the combination of two RNN models; backward and forward hidden layers to improve the performance of RNN neural network model. This approach can learn semantic information of words better because word semantics are correlated with preceding and succeeding information of the words.

D. Convolutional Neural Network

CNN is a multi-layer feed-forward neural network which improves the error in backpropagation network (BP) and reduces computation time and complexity of BP [49], [50]. It is recently used for sentiment classification because it can recognise local features by using convolution kernel, and automatically learns these features for classification solution.
CNN model consists of three main layers; convolution layer, pooling layer, and fully connected layer [51]. Sentences are converted into a matrix of numbers and input to the convolutional layer. Each sentence consists of words or tokens, and each token is corresponded to a row or vector on the matrix table. These vectors are typically generated by embedding techniques such as the Word2Vec and GloVe model.

CNN model takes the input of vectors and extracts local feature using filters. The most computations of features are performed in convolutional layer which is the most important layer in CNN. Convolutional layer produces feature maps using a function called convolution kernel.

After the convolution operation, pooling layer extracts the most important features. The pooling layer calculates local sufficient statistics. This process allows the pooling layer to reduce feature dimensions, makes CNN achieve computational time and cost reduction, and prevents the model from overfitting problem. Finally, the fully connected layer produces a probability distribution to classify sentiment results.

E. Long Short Term Memory

RNN is one of deep learning algorithms which is mainly used in NLP to predict the next word base on previously given words in a sentence. RNN also uses back-propagation as other traditional neural networks. However, RNN suffers from gradient exploding and vanishing problems. These two problems make RNN hard to train and fine-tune parameters. These problems normally occur during back-propagation process. Long Short Term Memory (LSTM) is an RNN model to improve the problems mentioned above.

LSTM modifies the structure of RNN. It reconstructs RNN layer into a structure that contains a gate and a memory unit. The purpose of LSTM is to keep the information in the memory cell for further utilisation and update. With this new structure, LSTM solves the problems of gradient exploding and vanishing problem in RNN. Moreover, it is more promising to apply LSTM to solve sentiment analysis problems because its variants can capture long short-term dependencies.

F. Bidirectional LSTM

Bi-LSTM is one of RNN algorithms to improve LSTM which has shortcomings of text sequence features. It solves the task of sequential modelling better than LSTM [52], [53]. In LSTM, information is flowed from backward to forward, whereas the information in Bi-LSTM flows in both directions; backward to forward and from forward to backward by using two hidden states. The structure of Bi-LSTM makes it a pioneer in sentiment classification because it can learn the context more effectively. Figure 2 shows the architecture of Bi-LSTM [54]. By utilising two ways of direction, input data of both preceding and succeeding sequence in Bi-LSTM are retained, unlike the standard RNN model that needs decay to include future data.

G. Sentiment Modalities

The objective of sentiment analysis is to uncover the expressed polarity within text, which involves interpreting opinions or emotions conveyed in both spoken and written language to determine the positive or negative sentiment. This analysis is particularly valuable in assessing the mood of stock investors. Sentiment analysis, a text mining and NLP method that examines subjective sentiments, holds significant importance in various domains such as product recommendations, healthcare, politics, and surveillance. In a forthcoming survey on predictive modelling using social media, Kalampokis et al. identified seven application areas covered by 52 published articles [55]. These areas include predicting or detecting disease outbreaks [56], election results [57], macroeconomic processes [58], box office performance of movies [59], natural phenomena like earthquakes [60], product sales [61], and financial markets [62]. The primary technique employed in these studies is sentiment analysis, where researchers perform semantic analysis on the contextual contents of each tweet to extract predictive insights from a selected group of individuals.

The study conducted by Wang et al. closely relates to the current research as they utilized tweets obtained from local news agencies [63]. They discovered initial evidence suggesting that these tweets have the potential to predict hit-and-run vehicular accidents and breaking-and-entering crimes. However, it is important to note that they only considered tweets from specific news agencies that were hand-selected. These tweets, authored by professional journalists, were relatively straightforward to analyse using existing text analysis techniques. However, this approach came at the cost of disregarding hundreds of thousands of potentially significant messages.

III. PROPOSED METHOD

Our objective is to leverage semantic knowledge learned in the text domain and historical crime data, and transfer it to a model trained crime prediction. We utilize our earlier developed model call ConvBiLSTM. The motivation of this model was to combine the strength from both CNN model and BiLSTM model. The study result of ConvBiLSTM

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3308967

proved that it could provide extra training by traversing the text twice from left to right and right to left, there by extracting the vector of the words in context of the information preceding and succeeding it and therefore can capture long term contextual dependencies and global features from the sequential text [14].

Our objective is to leverage the knowledge learn in the text domain from two pre-trained neural network models, and transfer it to a model trained for prediction crime. Figure 1illustrated the conceptual of the model. The word embedding vectors learn the language model from both crime and tweet.

The meaning of the tweet and crime data are summarized through convolution and pooling layer, and are constructed as a single representation of deep crime-semantic model by fusing the vector feature to capture the information from both crime and tweets data and train them to predict the crime.

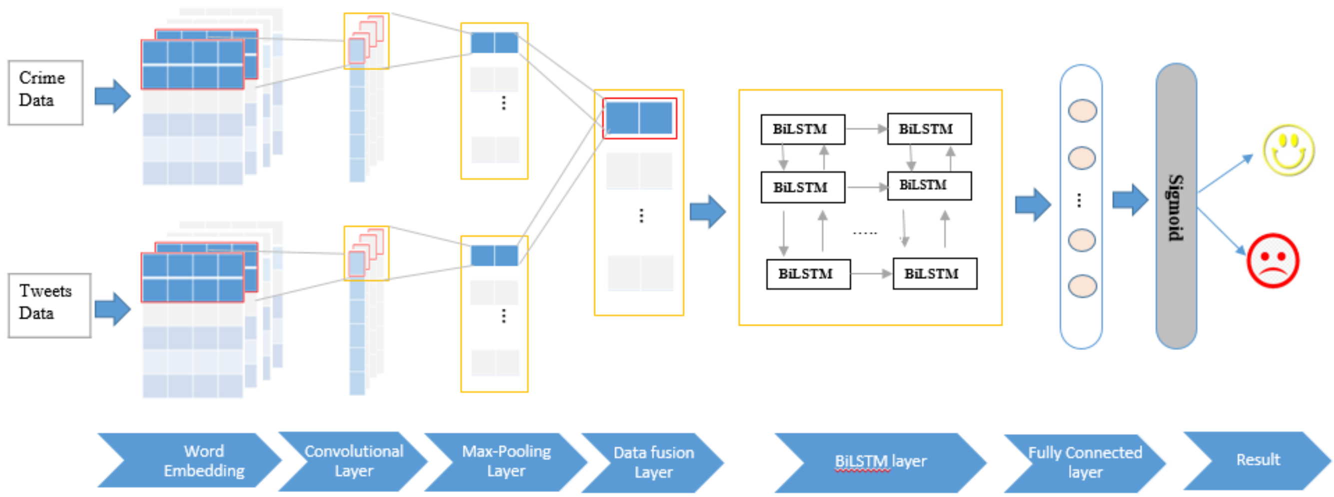

FIGURE 1. Conceptual Multimodal Data Fusion for Crime Prediction Using Tweets

A. WORD VECTORISATION

In this phase, the network takes the input of crime data and tweet data independently, and segments into word or token one by one. Each token is converted into a vector of numeric values. Word2Vec pre-trained word embedding models is used to generate word vector matrix. If each text of n words is represented as $T=\{w_1,w_2,...,w_n\}$, then each word is converted into word vector of d dimension, the text of input is defined as:

$$T=\{w_1,w_2,...,w_n\}\in R^{n*d} \tag{1}$$

Since the individual text of input have different lengths, its length needs to uniform (l). Its length was padded with zero-padding strategy. Text which has a length longer than the predefined length l will be truncated. But, if the text which has length shorter than l, zero padding will be added to the length. Therefore, all texts have the same dimension of matrix. Each text of l dimension is defined as follows:

$$T=\{w_1,w_2,...,w_n\}\in R^{l*d} \tag{2}$$

B. CONVOLUTIONAL LAYER

CNN model is good at extracting the most important words from tweets or sentences [39] and the convolution layer is the main step in CNN model. The word vectors matrix $T\in R^{l*d}$ from word embedding layer are fed into one-dimensional convolution layer. In one-dimensional convolution layer, the convolution word vector matrix is calculated through N filters and width q of convolution kernel to construct the local feature of n-gram. Filter F_n ,where $1\le n\le N$ generates feature maps as follows:

$$c_i^n = f(w^n \otimes X_{i:i+w-1} + b^n) \tag{3}$$

Weight matrix of filter F_n is defined as $w\in R^{q*d}$, and b^n is the bias of filter F_n, d is word vector dimension, and $\otimes$ is convolution operation, $X_{i:i+q-1}$ indicates that filter F_n extracts feature $X_{i:i+q-1}$ from X_i, f is non-linear activation, and the output of feature map of F_n filter is c_i^n where i th is element of c^n. In this study, RELU function was applied to non-linear activation f. For the sentence with length l, the following feature maps were obtained :

$$c=[c_1,c_2,...,c_i,c_l] \tag{4}$$

C. MAX-POOLING LAYER

Once convolution operation produces feature maps, pooling layer then extracts the most important features $\widehat{c}=\max\{c\}$ to calculate the local sufficient statistics. One-dimensional max-pooling converts each kernel size of input into a single

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3308967

output of the maximum number to reduce or down-sample version of the input. This is the reason why CNN model effectively reduces the number of features to prevent overfitting, also reduces time and complexity of parameters.

D. Data Fusion Layer

Feature level data fusion is applied in the study; it refers to the process of combining or integrating information from multiple sources or modalities at the level of extracted features. It involves extracting features independently from each modality and then merging them to create a unified feature representation that combines information from all sources. In feature-level fusion, the focus is on combining the extracted features rather than the raw data itself. The aim is to capture complementary information from different modalities or sources, creating a more comprehensive and informative representation for further analysis or decision-making. The fused data is computed as the average value of the individual data sources. The mathematical expression is:

Fused Data = (Data1 + Data2 + ... + Datan) / n (5)

where Data1, Data2, ..., Datan are the individual data sources, and n is the total number of data sources.

Before fusion, data pre-processing is conducted to ensure consistency and remove noise, including tokenization, removing stop words, stemming or lemmatization, and handling special characters or punctuation. Then, features are extracted from text data using word embedding; word2vec to capture the semantic meaning of words by representing them as dense vectors in a continuous space. The features are normalized to ensure compatibility and have the same scale range. Once the features are extracted and normalized, they can be fused to create a unified representation. Concatenation feature fusion is performed to concatenate into a single feature vector, combining the information from both crime and tweet features. The fused feature representation is then used as input for the desired analysis of crime prediction.

E. Bi-LSTM LAYER

In contrast to LSTM, Bi-LSTM allows the information to flow in both directions; backward to forward and from forward to backward by using two hidden states. This can help Bi-LSTM to learn the context better. By utilising these two-way directions, input data of both past and future information will be retained, whereby the standard RNN model needs decay to include future information. The principle implementation of Bi-LSTM is as; two opposite directions of LSTM network are connected to one output. The past information is obtained by forward LSTM state and the following information is obtained by backward LSTM state. This structure helps the network to retain preceding and succeeding information. The sequence output of the first layer in Bi-LSTM is the input of the second layer, and the sequence output of the second layer is the concatenation of the last unit output of forward and backward layers. After stacked Bi-LSTM layers, the final output is h :

$$h = [h_{forward}, h_{backward}] \tag{6}$$

F. DENSE LAYER AND RESULT

Dense layer is used in the model to connect each input with every output by using weights. Sigmoid is a function used in the final layer to produce the output. It takes the average of the random results into 1 and 0 forms. The prediction result of sigmoid function is presented in Equation (6). the result of sentiment is classified into either 0 or 1 by using binary cross-entropy. In this study, 0 represents a crime incident and 1 represents a non-crime incident.

$$Y = sigmoid(wh + b) \tag{7}$$

IV. EXPERIMENT

A. DATASETS

Two datasets were used in the study, first is the five crime types incident data including theft, battery, burglary, robbery, and motor vehicle theft. The second data is the tweet data associate with crime terminologies of the above crime. Crime incident dataset used is collected from Chicago city between 1-30 September 2019 via Chicago data web portal (https://data.cityofchicago.org/). The dataset consists of 22418 crime incidents.

Figure 2 shows the categories of past crime from the dataset. It can be seen that crimes are not equally distributed among categories but there is a huge difference between them. Theft crime is the most occurring crime happened in Chicago, with the incident number of more than 5000 (or 24.23%) follow by second most incident crime which is battery 19.46%.

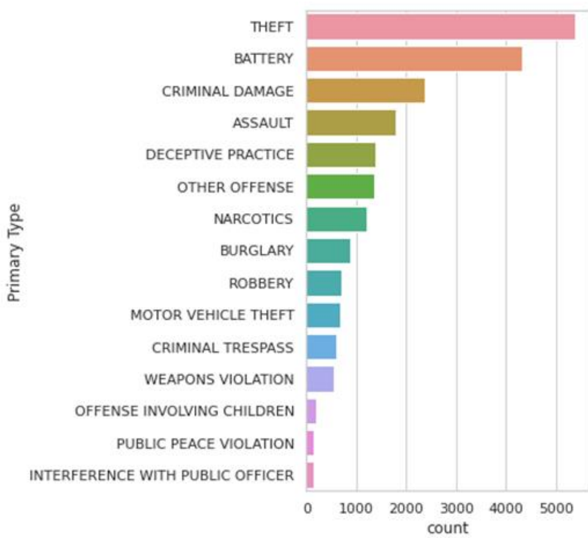

FIGURE 2. Categories of incident crime

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3308967

TABLE I
SAMPLE OF TWEETS

Date & Time	Tweets
2019-09-12 23:34:21	Employee theft, particularly at distribution canters, is a constant threat to supply chains. Wemobilize teams in escort vehicles for uninterrupted travel and constant surveillance. Our technology & expertise in freight security ensures that your supply chain is always protected
2019-09-12 13:25:35	Blue Star Security CEO Michael Sapraicone discusses the alarming and increasing trend of employee theft in his latest article
2019-09-24 15:54:34	Why aren't victims of theft also "workers"?
2019-09-23 14:34:10	Creighton got an absolute steal
2019-09-29 03:11:13	Haha. That's a good point. I'm gonna steal that.

Tweet dataset were collected from Twitter for the same window of time in Chicago, using python library called GetOldTweet. Total number of 398170 tweets were capture from the tweet. These tweets contain both crime terminology related public tweets and general public tweet in Chicago. Since the most past crime is theft and its similar crime types such as battery, burglary, robbery, and motor vehicle theft, we investigate these crime types in our crime prediction model. We investigated in detail to these crime types, the tweets which contain crime terminology such as theft and its synonym were extracted from the general tweets for the experiment. Crime terminology are including 'theft', 'crime', 'battery', 'steal', 'burglary', 'robbery', and "motor vehicle theft". A total number of 14,336 of tweets were extracted from the whole tweets. Table 1 shows the sample tweets collected from twitter.

B. DATA PRE-PROCESSING

Data pre-processing is the most crucial step in NLP because the raw dataset always consists of words or symbolst hat cannot be directly used by learning models. We started by removing @-mentions, Retweets represented as 'RT', links, and hashtags symbols using regular expressions as these things do not add any sort of value. A point to be noted here is that we made sure to not remove any words after the hashtag as it can contain a valuable reference to the sentiment of the tweet. For example: in #istandwith farmer, even though the symbol '#' does not add any positive or negative value to our analysis, the text "I standwith farmers" gives us insight into the state of the mind of the user. Further, the tweets containing special characters, punctuation, numbers, and emoticons were also removed.

Tokenization is defined as separating quantities of text into smaller units called tokens [28]. Tokenization is a fundamental step in modelling text data. It helps prepare the data before vectorization for understanding the meaning behind the text by analysing the sequence of the words. We used the porter stemmer method to reduce the inflection towards their root forms. This was done by stripping the suffix to produce stems [29]. Lastly, the fully pre-processed tweets were stored in a new pandas column called "stem_tweets" in our existing data frame of tweets and crime datasets.

D. HYPER-PARAMETERS SETTING

In many cases, the model may produce less accuracy or even produce overfitting or under-fitting. To obtain high model performance, conducting hyper-parameters tuning is very critical. Therefore, the randomised search strategy was used to tune hyper-parameter and optimise the accuracy. Table 2 describes the hyper-parameters value in the proposed model.

TABLE II
HYPER-PARAMETERS SETTING

Parameters	Values
Embedding dimension	Word2Vec =300
Kernel size	5
Filter	128
Pool size	2
Bi-LSTM output size	64
Kernel regularization	L2(0.001)
Weight constraints	Kernel constraint (max_norm=3)
Activation	Relu
(Recurrent) dropout	0.1
Batch size	128
Batch normalization	yes
Loss function	Cross-entropy
Optimizer	Adam
Learning rate	0.001

V. RESULT AND DISCUSSION

This section covers the visualization and analysis result of the model. Figure 3 shows the word cloud obtained from the cleaned tweets. Word Cloud is a pictorial representation of commonly used words in a particular dataset. We provided our dataset of cleaned tweets to the model to generate this word cloud. The entire word cloud represents the most frequently used words. The words with a larger font occur more commonly than the words with a smaller font. The word cloud can give us an overview of the theft crime. It can also help us in understanding the essence of the crime. In our word cloud of cleaned tweets, the words that occur most frequently are grand theft, employee theft, cargo theft, high crime, theft auto and many others. While the words like theft bring to light the common motive of the tweets, the words like cargo and employee indicate that people were tweeting in expressing the theft happened related to cargo and employee. A copious number of words like steal, business damage, theft occurs, chance theft and theft spell have been

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3308967

mentioned which reveal that Twitter users expressed the concern of increasing theft.

FIGURE 3. Word Cloud of the Tweets

A. Baseline Comparison

To evaluate the performance of our multimodal deep learning for crime prediction, we employed some existing methods for crime prediction comparison:

- SVM: conduct a crime prediction on crime data set, without considering tweet data.
- Logistic Regression: apply crime data set to prediction crime data, without considering tweet data
- Sentiment Based SVM: to prediction crime data, sentiment score of the associate tweet were added to crime data.
- Sentiment based Logistic Regression: sentiment score of the associated tweet were added to crime data set for crime prediction.
- NAHC: neural attentive framework for hour-level crime prediction to address the challenge at day level. The framework integrates Gated Recurrent Units (GRUs) with a temporal attention mechanism in order to capture both short-term and long-term temporal relationships, while also considering time-sensitive external factors [64].
- DNN with feature level data fusion: crime occurrence prediction model incorporates environmental context information through the fusion of multi-modal data [65].
- CrimeTelescope: a platform for online crime prediction and visualization that utilizes the fusion of urban and social media data [66].
- ANN+BERT: BERT base approach to detect the crime related twitter post [67].
- BERT base Model: A crime detection model base on crime related posts from twitter [68].

B. Result and Discussion

The test accuracy results of our model and other methods for crime prediction are presented in Table 3. Upon analysing the table, we observe that sentiment-based crime prediction models, such as sentiment-based SVM and sentiment-based logistic regression, outperform SVM and logistic regression models that do not consider sentiment data. This indicates that incorporating tweet sentiment into crime prediction models significantly enhances their performance compared to traditional models that do not consider tweet sentiment. Therefore, it can be concluded that tweet sentiment has a noticeable impact on the outcome of crime prediction.

TABLE III
RESULT COMPARISON OF CRIME PREDICTION MODEL WITH OTHER APPROACHES

Model	*Task*	*Dataset*	*Accuracy (%)*
SVM	Crime prediction model base on crime historical data	Crime data in Chicago, September 2019	78.45%
Sentiment based SVM	Crime prediction model using crime data and tweets	-Crime data in Chicago, September 2019 - Tweets in Chicago	79.18%
Logistic Regression	Crime prediction model base on crime historical data	Crime data in Chicago, September 2019	86.37%
Sentiment based Logistic Regression	Crime prediction model using crime data and tweets	- Crime data in Chicago, September 2019 - Tweets in Chicago	86.94%
NAHC	Neural attentive framework for hour-level crime prediction to address the challenge at day level.	- Crime data from Xiaogan, China - Crime data from NYC - POIs data - Meteorological data	61%
DNN with feature level fusion	Crime prediction model incorporates environmental context information through the fusion of multi-modal data	- Crime data from Chicago - Environmental data - Spatial data	84.25%

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3308967

CrimeTelescope	Online platform for crime prediction and visualization that combines features from various sources	- Crime data from NYC - Tweets data - Urban Infrastructure data-POIs	80%
ANN+BERT	BERT-based approach to detect crime using tweets and weather data	- Crime related tweets - Weather data	91.5%
BERT-based model	Crime detection, which utilizes BERT model to detect crime-related posts from Twitter	- Crime related Tweets	92.8%
Multimodal data fusion for crime prediction (Proposed Model)	Crime prediction using data fusion and ConvBiLSTM	- Crime data in Chicago, September 2019 - Tweets Data	97.75%

Furthermore, our multimodal data fusion approach for crime prediction yielded the highest result, showcasing an improvement over sentiment-based logistic regression and over sentiment-based SVM. There are two reasons why our model achieved such results. Firstly, both sentiment-based SVM and sentiment-based logistic regression models solely consider the sentiment polarity and directly add sentiment features to the crime data. However, these features do not convey the same sentiment-driven information, limiting their effectiveness.

In comparison to NAHC [64], a neural attentive model introduced for hour-level crime prediction, our model achieves a higher accuracy rate of 36.75%. NAHC utilizes four datasets, including crime statistics from the police department of Xiaogan (China), crime statistics from the police department of NYC, POI data, and meteorological data. However, its accuracy is only 61%. This discrepancy can be attributed to NAHC's lack of consideration for tweet data and its utilization of multi-graph convolution instead of a data fusion approach.

In comparison to DNN with feature level fusion model [65] which presents a well-designed data fusion approach, where multiple datasets are fused at the feature level. The model combines crime temporal, environmental (image), and spatial (demographics) features. Lower accuracy highlights the importance of tweet data in crime prediction, which the DNN with feature level fusion model fails to consider. Additionally, the DNN model suffers from imbalanced data due to the overwhelming lack of crime occurrence reports across all sampling points of the environmental feature.

In comparison to CrimeTelescope [66] which is an online platform for crime prediction and visualization that combines features from various sources, such as crime temporal data, tweets, and points of interest (POI) in urban areas. While the model's data fusion design is commendable, the discrepancy in accuracy can be attributed to the fact that CrimeTelescope employs Latent Dirichlet Allocation (LDA) for tweet feature extraction, whereas our model utilizes Word2Vec.

When compared to the ANN+BERT model [67], which employs a BERT-based approach to detect crime using tweets and weather data, we achieved an higher accuracy rate. Although the ANN + BERT model utilizes tweets for crime prediction, it employs a direct concatenation method on the data source, which can lead to issues such as overfitting, redundancy, and dependency on multiple datasets.

Our model surpasses the BERT-based model in crime detection, which utilizes crime-related posts from Twitter. This finding aligns with the approach employed in the ANN + BERT model described in [68], where a direct concatenation method is implemented.

Therefore, our ConvBiLSTM model, which incorporates multimodal data fusion, demonstrates superior performance compared to other models, including SVM, Logistic Regression, NAHC, DNN with feature-level data fusion, CrimeTelescope, ANN+BERT, and BERT-based crime prediction models. By utilizing ConvBiLSTM, the network benefits from both forward and backward LSTM hidden layers, enabling a more comprehensive context understanding for the output layer. As a result, our multimodal data fusion approach, utilizing tweet data and ConvBiLSTM, outperforms traditional deep-learning and BERT models.

V. CONCLUSION

We have introduced multimodal deep learning modal for crime prediction, which can nicely combine the information of tweets and crime incident data. The proposed multimodal deep learning relies on the combination architecture of CNN and BiLSTM, called ConvBiLSTM. We modified the structure of ConvBiLSTM; pre-train crime modal and tweet model independently and parallel, summarize the meaning of both modals through convolution and pooling layer, construct a single representation of deep crime-semantic model by fusing the vector feature to capture the information from both crime and tweets data and train them to predict the crime.

Our approach features a novel ranking model that aligned parts of language modalities through a common, multimodal embedding. We showed that this model provides state of the art performance on crime prediction using tweets. Second, we described a Multimodal ConvBiLSTM that can fuse data from crime modal and tweet modal. We evaluated its performance against other baseline models.

The analysis in this study was limited to tweets written in English and related to theft crime and its similar crime such as battery, burglary, robbery, and motor vehicle theft. Future studies can expand the analysis into different languages and different crime types. Furthermore, the study is conducted for only one month of window time on database, an expansion of window time could help verifying the result of the analysis. In addition, the findings of this study are limited to only users on Twitter platform; future research can explore text content from other social platform to compare the results. Future research may also look at various algorithms including using supervised and unsupervised learning as method and see if the outcomes vary from this study.

REFERENCES

[1] Sathyadevan, Shiju& S., Devan &Gangadharan, Surya. (2014). Crime Analysis and Prediction Using Data Mining. 10.1109/CNSC.2014.6906719.

[2] B. Liu, “Sentiment analysis: Mining opinions, sentiments, and emotions,” *Cambridge University Press*, 2015.

[3] B. Liu, “Sentiment analysis and opinion mining, Synthesis Lectures on Human Language Technologies,*” Morgan & Claypool Publishers*, vol. 5, no. 1, pp. 1-167, 2012, Accessed: Nov. 1, 2020 [Online]. Available: https://doi.org/10.2200/S00416ED1V01Y201204HLT016.

[4] B. Pang and L. Lee, “A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts,” *In Proceedings of the 42nd annual meeting on association for computational linguistics*, pp. 271-278, 2004.

[5] P. D. Turney, “Thumbs up or thumbs down: semantic orientation applied to unsupervised classification of reviews.” *In Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 417- 424, 2002.

[6] Bankmycell. “How Many Users Does Twitter Have? (May 2023)” Accessed: May. 20, 2023 [Online]. Available: https://www.bankmycell.com/blog/how-many-users-does-twitter-have

[7] A. Salim and E. Omer, “Cybercrime Profiling: Text mining techniques to detect and predict criminal activities in microblog posts*,”International Conference on Intelligent Systems: Theories and Applications (SITA)*, pp. 1-5, 2015.

[8] B. Liu, “Sentiment Analysis: Mining opinions, sentiments, and emotions,” *Computer Linguistics*, vol. 42, no. 3, pp. 1-4, 2016.

[9] D. Lahat, T. Adali and C. Jutten, "Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects," in Proceedings of the IEEE, vol. 103, no. 9, pp. 1449-1477, Sept. 2015, doi: 10.1109/JPROC.2015.2460697.

[10] M. S. Gerber, “Predicting crime using Twitter and kernel density estimation”, Decision Support System, vol. 61, no.1, pp. 115–125, 2014.

[11] Hall DL, Llinas J. An introduction to multisensor data fusion. Proc IEEE 1997;85(1):6–23.

[12] Durrant-Whyte HF. Sensor models and multisensor integration. Int J Robot Res 1988;7:97–113.

[13] Castanedo F. A review of data fusion techniques. Sci World J2013;2013:704504.

[14] S. Tam, R. B. Said and Ö. Ö. Tanriöver, "A ConvBiLSTM Deep Learning Model-Based Approach for Twitter Sentiment Classification," in IEEE Access, vol. 9, pp. 41283-41293, 2021, doi: 10.1109/ACCESS.2021.3064830.

[15] H. McGurk and J. MacDonald, “Hearing lips and seeing voices,” Nature, vol. 264, no. 5588, p. 746, 1976

[16] B. P. Yuhas, M. H. Goldstein, and T. J. Sejnowski, “Integration of acoustic and visual speech signals using neural networks,” IEEE Communications Magazine, vol. 27, no. 11, pp. 65–71, 1989.

[17] H. Bourlard and S. Dupont, “A new asr approach based on independent processing and recombination of partial frequency bands,” in Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on, 1996, pp. 426C–429.

[18] B. Matthew, “Coupled hidden markov models for complex action recognition,” in Proceedings of Computer Vision and Pattern Recognition, 1997, pp. 994C–999.

[19] C. G. M. Snoek and M. Worring, “Multimodal video indexing: A review of the state-of-the-art,” Multimedia Tools & Applications, vol. 25, no. 1, pp. 5–35, 2005.

[20] S. K. D’Mello and J. Kory, “A review and meta-analysis of multimodal affect detection systems,” ACM Computing Surveys, vol. 47, no. 3, pp.1–36, 2015.

[21] T. Baltruˇsaitis, C. Ahuja, and L.-P. Morency, “Multimodal machine learning: A survey and taxonomy,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443, 2019

[22] J. Cheng, Y. Dai, Y. Yuan and H. Zhu, "A Simple Analysis of Multimodal Data Fusion," 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Guangzhou, China, 2020, pp. 1472-1475, doi: 10.1109/TrustCom50675.2020.00199.

[23] C. Shao, H. Li, and L. Ma, “Visual cognitive mechanism guided video shot segmentation,” in Cognitive Computing – ICCC 2019, R. Xu, J. Wang, and L.-J. Zhang, Eds. Cham: Springer International Publishing, 2019, pp. 186–196.

[24] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell, “Long-term recurrent convolutional networks for visual recognition and description,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 677–691, 2017.

[25] X.-Y. Wu, C.-N. Gu, and S.-J. Wang, “Special video classification based on multitask learning and

multimodal feature fusion," Optics and Precision Engineering, vol. 28, no. 5, pp. 1177–1186, 7 2020.

[26] Y. Dai, G. Wang, and K.-C. Li, "Conceptual alignment deep neural networks," Journal of Intelligent & Fuzzy Systems, vol. 34, no. 3, pp.1631–1642, 2018.

[27] Y. Dai, G. Wang, J. Dai, and O. Geman, "A multimodal deep architecture for traditional chinese medicine diagnosis," Concurrency and Computation: Practice and Experience, vol. In Press, 2020. [Online]. Available: https://dx.doi.org/10.1002/cpe.5781

[28] E. Shutova, D. Kiela, and J. Maillard, "Black holes and white rabbits: Metaphor identification with visual features," in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, pp. 160–170. [Online]. Available: https://www.aclweb.org/anthology/N16- 1020

[29] G. Evangelopoulos, A. Zlatintsi, A. Potamianos, and P. Maragos, "Multimodal saliency and fusion for movie summarization based on aural, visual, and textual attention," IEEE Transactions on Multimedia, vol. 15, no. 7, pp. 1553–1568, 2013.

[30] E. Morvant, A. Habrard, and S. Ayache, "Majority vote of diverse classifiers for late fusion," in Structural, Syntactic, and Statistical Pattern Recognition, 2014, pp. 153–162

[31] M. S. Gerber, "Predicting crime using Twitter and kernel density estimation", Decision Support System, vol. 61, no.1, pp. 115–125, 2014.

[32] X. Wang, M.S. Gerber and D.E. Brown, "Automatic crime prediction using events extracted from twitter posts, in Social Computing, Behavioral-Cultural Modeling and Prediction,". Springer, 2012, pp. 231–238.

[33] X. Chen, Y. Cho and S.Y. Jang," Crime prediction using twitter and weather," Systems and Information Engineering Design Symposium (SIEDS), IEEE, 2015, pp. 63–68.

[34] T. Brandt, J. Bendler, and D. Neumann, "Infomation& Management Social media analytics and value creation in urban smart tourism ecosystems," Information and Management., vol. 54, no. 6, pp. 703–713, 2017.

[35] N. Malleson and M.A. Andresen, "The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns," Cartography and Geographic Information Science, 42(2), pp.112–121, 2015.

[36] N. Zainuddin, A. Selamat, and R. Ibrahim, "Improving Twitter AspectBased Sentiment Analysis Using Hybrid Approach," Intelligent Information and Database Systems, vol. 9621, N. T. Nguyen, B. Trawiński, H. Fujita, and T.-P. Hong, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 151–160.

[37] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learn-ing techniques," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing vol 10, 2002, pp. 79–86.

[38] F. Y. Zhou, L. P. Jin and J. Dong, "Review of convolutional neural network," *Chin. J. Comput. Vol. 1,* pp. 35-38, 2017

[39] Y. Li and H. B. Dong, "Text emotion analysis based on CNN and BiLSTM network feature fusion," *Comput. Appl, vol 38, no. 11,* pp. 29-34, 2018.

[40] N. Kalchbrenner and P. Blunsom, "Recurrent convolutional neural networks for discourse compositionality," *Comput. Sci, vol. 10,* pp. 1-2, 2013.

[41] K. Kim, B. S. Chung and Y. R. Choi, "Language independent semantic kernels for short-text classification," Expert Syst. *Appl. Int. J. vol. 41, no. 2,* pp. 735-743, 2014.

[42] S. Liu, "Novel unequal clustering routing protocol considering based on network partition & distance for mobile education," *J. Netw. Comput. Appl. Vol. 88, no. 15,* pp. 1-9, 2017.

[43] S. Zhou, "A low duty cycle efficient MAC protocol based on self-adaption and predictive strategy" *Mob. Netw. Appl, vol. 23, no. 4,* pp. 828-839, 2018.

[44] C. Jin and W. Li, "Chinese word segmentation based on bidirectional LSTM neural network model," *Chin. J. Inform, vol. 32, no. 2,* pp. 29-37, 2018.

[45] J. Chen, H. F. Li and L. Ma, "Dimensional speech emotion recognition method based on multi-granularity feature fusion," *Signal Process, vol. 33, no. 3,* pp. 374-382, 2017.

[46] D. G. Zhang, H. L. Niu and S. Liu, "Novel PEECR-based clustering Routing approach," *Soft. Comput, vol. 21, no. 24,* pp. 7313-7323, 2017.

[47] Y. M. Tang, "Novel reliable routing method for engineering of internet of vehicles based on graph theory," *Eng. Comput, vol. 36, no. 1,* pp. 226-247, 2019,

[48] Y. X. Fan, J. F. Guo and Y. Y. Lan, "Context-based deep semantic sentence retrieval model," *Chin. J. Inform. Sci, vol. 31, no. 5, pp.* 161-167, 2017.

[49] K. Simonyan and A. Zisserman, "Two-Stream Convolutional Networks for Action Recognition in Videos," University of Oxford, 2014.

[50] F. Chollet, "Deep Learning with Python," Shelter, Island: Manning, 2017.

[51] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: A review," Artif. Intell. Rev., vol. 151, 2019, DOI: 10.1007/s10462-019-09794-5.

[52] W. Liu, P. Liu, Y. Yang, Y. Gao and Y. Yi, "An attention-based syntax-tree and tree-LSTM model for sentence summarization," Int. J. Performab. Eng, vol. 13 no. 5, pp. 775-782, 2017.

[53] X. Niu, Y. Hou and P. Wang, "Bi-directional LSTM with quantum attention mechanism for sentence modelling, in: Proceedings of the 24th International Conference on Neural Information Processing," Springer Verlag, Guangzhou, China, pp. 178-188, 2017.

[54] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2673-2681, 1997.
[55] Kalampokis, E,.Tambouris, E,.&Tarabanis, K. (2013). Understanding the predictive power of social media, Internet Research, 23.
[56] Culotta, A,.& Huberman, B. (2010). Towards detecting influenza epidemics by analyzing Twitter messages, Proceedings of the First Workshop on Social Media Analytics, ACM, 115–122.
[57] Franch, F. (2013). Wisdom of the crowds 2: 2010 UK election prediction with social media, Journal of Information Technology & Politics, 10, 57–71.
[58] Velay, M., and Daniel, F. (2018). Using NLP on news headlines to predict index trends. Retrieved from http://arxiv.org/abs/1806.09533
[59] Asur, S,.& Huberman, B. (2010). Predicting the future with social media, IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, IEEE, 492–499.
[60] Earle, P, S,. Bowden, D, C,.& Guy, M. (2012). Twitter earthquake detection: earthquake monitoring in a social world, Annals of Geophysics, 54.
[61] Choi, H,.& Varian, H. (2012). Predicting the present with Google Trends, The Economic Record, 88, 2–9.
[62] Bollen, J,. Mao, H,.& Zeng, X. (2011). Twitter mood predicts the stock market, Journal of Computational Science, 2, 1–8.
[63] Wang, X., Brown, D,. & Gerber, M. (2012). Spatio-temporal modelling of criminal incidents using geographic, demographic, and Twitter-derived information, Intelligence and Security Informatics. Lecture Notes in Computer Science IEEE Press.
[64] Liang. W., Wang. H., Tao. H.& Cao. J., (2022), Towards hour-level crime prediction: A neural attentive framework with spatial–temporal-categorical fusion, Neurocomputing, 486, pp. 286–297, DOI: https://doi.org/10.1016/j.neucom.2021.11.052
[65] Kang. H.W.,& Kang. H.B., (2017), Prediction of crime occurrence form multi-modal data using deep learning,PLoS ONE, Vol.12, No.4, DOI: https://doi.org/10.1371/journal.pone.0176244
[66] Yang. D., Heaney. T., Tonon, Wang. L. &Cudre-Mauroux. P., (2018),CrimeTelescope: crime hotspot prediction based on urban and social media data fusion. World Wide Web vol.21, pp.1323–1347, DOI: https://doi.org/10.1007/s11280-017-0515-4
[67] S. P. C. W. Sandagiri, B. T. G. S. Kumara and B. Kuhaneswaran, "ANN Based Crime Detection and Prediction using Twitter Posts and Weather Data," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), Sakheer, Bahrain, 2020, pp. 1-5, doi: 10.1109/ICDABI51230.2020.9325660.
[68] S. P. C. W. Sandagiri, B. T. G. S. Kumara and B. Kuhaneswaran, "Deep Neural Network-Based Approach to Identify the Crime Related Twitter Posts," 2020 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 2020, pp. 1000-1004, doi: 10.1109/DASA51403.2020.9317098.

SAKIRIN TAM was born in Kompong Cham, Cambodia, in 1985. He received the B.S. degree in information technology from Fatoni University, Pattani, Thailand, in 2010 and the M.S. degree in information technology management from University Technology Malaysia, Johor Bahru, Malaysia, in 2014. He is currently vice dean of ICT faculty at University of Puthisastra, Cambodia. From 2010 to 2015, he was a lecture at Information Technology Department, Fatoni University, Pattani, Thailand. His research interest includes machine learning, deep leaning, natural language processing, sentiment analysis, data mining, green IT and computer based test.

Ö. ÖZGÜR TANRIÖVER holds Computer Engineering BSc degree and Science and Technology Policy MSc and Information Systems Ph.D. degrees from Middle East Technical University, Ankara, Turkey. He was a certified Information Systems Auditor (CISA) in the Information Management Department at the Banking Regulation Agency. He is currently an Assistant Professor in the Computer Engineering Department of Ankara University. His current research interests are software quality, information systems security and machine learning for prediction.